

Academy development signals a gradual leap forward

China's scientists are highly valued but poorly paid, and great in number but small in productivity. A new initiative by the Chinese Academy of Sciences should improve matters.

Despite its increasing acceptance of market thinking, China remains a country that develops above all by planning. But what actually happens there depends on regional influences as much as on the central government, and on initiatives and bartering by ministries and agencies as much as on any decision by a prime minister or president. In particular, China's science policy can sometimes be like the thinking behind its Internet development: opaque to the Chinese community and the outside world, hindered by lack of coordination and by the self-interest of the various owners of different portions, but in the long run successful in delivering improved resources and quality. But significant progress has to be measured over periods as long as a decade.

It is not easy, therefore, to be sure when something significant is, in fact, happening. The plan announced last week to revamp the research activities of the Chinese Academy of Sciences (CAS) surprised some senior figures in other parts of the research policy structure by the sheer scale of what is being proposed (see page 7). Large quantities of additional funds are being mentioned, but the negotiations between the academy and the finance authorities are yet to be concluded. Nevertheless, from amid much rhetoric about the importance of science and high technology to the future of China, this is one of the most significant developments that can be expected over the next few years.

Although there have been substantial improvements in university science departments, the academy of sciences still remains an important source of much of China's best science. But with only 14,459 papers published by the whole of China in 1996 in the Science Citation Index, one is left wondering what the majority of the 68,000 researchers in the academy do with their time.

Scientists by and large accept that the government needs to focus on the economic and social needs of what is after all a 'developing

country'—basic research that has no strategic relevance will be very low on the agenda for the foreseeable future. The emphasis in the initial research priorities of the academy's plan is accordingly very much on relevant science. But China's researchers are less sanguine about their poor ability to retain younger scientists because of pitifully low salaries, and the lack of support for software and consumables: the combination means that there is rather too much good equipment sitting around underused. The academy's intention to boost the salaries of new appointments with bonuses, especially for those enticed home from abroad, is to be welcomed, but will not significantly improve the ability to attract, let alone retain, young talent from those educated and trained at home, where people can earn more money as a chauffeur than as a scientist.

One essential way forward is for the academy to find a way of reducing the number of staff that it has on its books. The academy's plan does call for more than a halving in permanent staff, but not until the end of the next decade, which suggests that the plan relies heavily on retirements rather than on tough decisions to 'redirect' unproductive personnel.

Despite its difficulties, the academy—like other parts of the system—has recognized in recent years the importance of internationalism. Gone are the days when, in a spirit of misplaced patriotism, scientists were required to publish in Chinese journals at the expense of international recognition. If anything, the pendulum may be swinging too vigorously in the opposite direction, with strong incentives being offered for publication in top journals to an extent that may ultimately be demotivating, given the realities of what can be achieved (though the best is indeed of that quality). But an increasing acceptance of international judgements, in appraisals of institutions as well as individuals, is to be strongly encouraged, and will help ensure that new developments are built on solid ground. □

Environment agency needs more tact

The goodwill that greeted the appointment of UNEP's new executive director is in danger of evaporating.

In five months as executive director of the United Nations Environment Programme (UNEP), Klaus Töpfer has managed what his predecessor took several years to achieve: he has provoked the ire of many parties to the two main UN environment conventions, as well as the head of the UN's main environment funding mechanism (see page 4). This is in stark contrast to the immense goodwill that greeted his appointment. What has gone wrong?

To some extent, Töpfer faced an intractable task. He was appointed to revive a flagging UN agency, many of whose own governing council of ministers would be happy to keep things as they are. Developing countries, in particular, are reluctant for UNEP to interfere with the administration of existing UN environment conventions. They are also a little concerned at the new executive director's tendency to think aloud and in public, rather than in the UN's notorious corridors and

back rooms. Not surprisingly, developing countries have become some of Töpfer's strongest—if covert—opponents.

But Töpfer is right to seek better coordination between the UN's unwieldy network of environment conventions, and better use of UNEP's impressive access to scientific expertise. And, in line with trends in European countries, he wants a more formal role for non-government groups in the advisory process. At the same time, he cannot afford to risk alienating those whose support he needs to carry the reforms through.

Having refreshingly made suggestions for change, and chaired a task force on reforming the handling of environmental issues across the UN, it is time for Töpfer to take a more considered approach, with a more sympathetic understanding of the complexities of North-South politics. He has time on his side. □



Japan reacts uneasily to nuclear plan for meeting carbon cuts

[TOKYO] Tension is growing in Japan over the latest detailed proposals to tackle global warming, which call for the active promotion of nuclear energy to help reduce the country's emissions of greenhouse gases.

The general policy outlines released last week by a task force on global warming, representing nine government ministries and agencies, are intended to provide specific measures for implementing Japan's commitment at last year's climate conference in Kyoto. Japan is committed to curbing emissions of greenhouse gases by 6 per cent from 1990 levels between 2008 and 2010.

The proposals say that Japan's output of nuclear energy needs to be increased substantially to implement such a reduction, and that the total amount of electricity generated by nuclear power will need to grow by 50 per cent from its 1997 level by 2010.

But the electricity industry says that such an increase would be impossible, given the current antinuclear mood in Japan. According to officials at the Ministry of Trade and Industry (MITI), 20 nuclear reactors would have to be built to meet the targets.

Under the working group's proposals, Japan will aim to reduce emissions of carbon dioxide, nitrogen dioxide and methane by 2.5 per cent from 1990 levels by 2010, and to restrict the increase of emissions of hydrofluorocarbons, perfluorocarbons and sulphur hexafluoride to 2 per cent.

Although the government has previously raised the possibility of achieving a 3.7 per cent reduction by including 'sink gases' — the amount of carbon dioxide absorbed by forests — the working group warns against relying on such a method. It argues that techniques for counting sink gases remain unproven, and that Japan will be expected to reduce its emissions by a further 3.7 per cent if such a method is adopted.

There is concern about the reliance on nuclear power because a series of accidents and subsequent cover-ups at nuclear facilities run by the Power Reactor and Nuclear Fuel Development Corporation (PNC) have led the public to lose confidence in Japan's nuclear policy (see *Nature* **386**, 209; 1997).

"It would be extremely difficult to build more nuclear reactors in the current climate, considering how sensitive people are about nuclear safety," says a director at Tohoku Electric Power, which had to freeze its plans to build nuclear power reactors in Niigata Prefecture, in the north of Japan, in the face of strong opposition from local residents.

Tokyo Electric Power, one of the world's largest private electricity companies, faced



Waste not, want not: but nuclear facilities such as Rokkasho's reprocessing plant are out of favour.

similar problems when trying to build two reactors at its site in Fukushima, in north-east Japan. The company, which depends on nuclear energy for 40 per cent of its output, has no immediate plans to build more reactors; instead it will concentrate on improving the efficiency of its thermal power plants.

The electricity industry is facing several problems related to Japan's plans to curb the emission of greenhouse gases. For example, although it has projected a 50 per cent expansion in demand for power between 1990 and 2010, it is facing a target, set by MITI, to limit its increase in emissions to less than 20 per

cent during that same period.

But the biggest hurdle so far is the plan to promote nuclear energy, which is unlikely to be implemented unless the central government agrees to give the industry its active support for such projects.

Mitsuo Horiuchi, the trade minister, said at a press conference last week that a major effort will be made to secure public acceptance of new nuclear power facilities, and that the question of radioactive waste disposal will be looked into urgently.

Another important task would be regaining public confidence. A white paper on nuclear energy issued last week by the Science and Technology Agency attributes public distrust in the government's nuclear policy to the discrepancy between 'technological safety' as seen by the experts and the 'sense of security' in the general public.

The report promises to restore the nation's trust in nuclear safety, and to endorse the importance of nuclear power as a future energy resource.

PNC, which is set to be transferred to a new semi-public organization in October, is likely to be given responsibility for handling the issue of radioactive waste disposal, which it calculates is likely to start between 2030 and 2040.

Asako Saegusa

'Ignition not vital' in scaled-down fusion reactor

[TOKYO] The International Thermonuclear Experimental Reactor (ITER), a proposed US\$10 billion project to build a next-generation fusion facility, faces significant scaling down after its council last week adopted new design parameters for a smaller, cheaper version.

At a meeting in Tokyo, council members of the ITER partners — the United States, Europe, Japan and Russia — approved proposals by its working group to reduce costs by modifying some of its technical objectives.

The working group, set up in February to explore less expensive designs for the reactor, proposed in May that the cost could be cut by up to half by reducing the radius of the doughnut-shaped confinement vessel

from 8 metres to 6 metres (see *Nature* **393**, 406; 1998), and by reducing its output from 1.5 GW to 0.7 GW.

Although the new design would alleviate budgetary constraints, the reactor's ignition — required to generate a self-sustaining thermonuclear burn — is likely to be lost. But officials from the Japanese Atomic Energy Research Institute (JAERI), which put forward the plans for the scaled-down version, say that ignition may not be essential for carrying out successful nuclear fusion.

"Although some of ITER's aims have been relaxed, the proposed changes are not viewed as a compromise," says Hiroshi Kishimoto, an executive director.

Japan is the most likely candidate to host the site, as

it is expected to put up the largest share of the cost for the international project.

At a meeting earlier this month, Japan's Atomic Energy Council decided to present ITER's design team with detailed data on Japan's regional characteristics, such as information on earthquake prevalence and maps of coastal areas adequate for placing fusion reactors.

Although it declined to look into sites, the council created a subcommittee on fusion energy development that will discuss Japan's future effort in nuclear-fusion research and the possibility of collaboration with industry. But the government says it will not make any decision about ITER's location for two years because of the nation's current economic crisis. **A.S.**

UNEP rebuffed in search for a wider role

[LONDON] Suggestions for increasing the influence of the United Nations Environment Programme (UNEP) have run into strong opposition from both developing countries and the Global Environment Facility (GEF), which finances environmental projects.

GEF has objected vigorously to a proposal by Klaus Töpfer, UNEP's executive director, that UNEP should play a stronger role in shaping GEF priorities and programmes. The proposal was contained in a draft report by a high-level task force on reforming environment policy across the UN system, set up by UN Secretary General Kofi Annan and chaired by Töpfer.

In an unusually strongly worded letter, Mohamed El-Ashry, GEF's chief executive, has warned Töpfer that UNEP "might find itself on a collision course with the GEF Council" unless "some inaccurate references to the GEF" were changed. The changes, primarily concerning UNEP's role with respect to GEF, are understood to have been made.

El-Ashry wrote that the fund's governing council was "the only entity in the GEF structure" with responsibility for guiding funding. UNEP, he pointed out, was among GEF's "implementing agencies", and therefore had to be accountable to the GEF council. Other implementing agencies are the UN Development Programme and the World Bank.

This is not the first time Töpfer has met opposition to his efforts to find a new role for UNEP. Last month, developing countries



Töpfer: committed to a stronger UNEP.

blocked his suggestion to set up an intergovernmental panel of experts to explore the use of economic instruments — such as a greenhouse-gas trading system to reduce emissions — to protect the environment.

Töpfer had floated this idea at a meeting of the UN climate convention's advisory body of scientific experts in Bonn. But scientists representing the Group of 77 developing countries said research into economic instruments was already dealt with by one of the working groups of the Intergovernmental Panel on Climate Change (IPCC).

Developing countries were similarly critical of an earlier Töpfer suggestion, delivered to the conference of the UN biodiversity convention in Bratislava, that UNEP could take over much of the convention's scientific work (see *Nature* 393, 99; 1998).

Töpfer now faces a dilemma. Few of his critics disagree with the need to strengthen environment policy across the UN system, but few are convinced that this should be done through UNEP, which they believe has outlived its usefulness as most of the environmental conventions and institutions it has spawned are now accountable to their

own member countries.

Indeed, many countries now see greater involvement by UNEP as a form of interference. But Töpfer, an experienced German politician with more than a decade spent at cabinet level, says he remains committed to his goal of strengthening UNEP.

To that end, he repeats his desire to see non-government groups play a greater role in the decision-making process, and feels strongly that UNEP should take more responsibility for scientific assessments and environmental monitoring. "These are not just my suggestions. They are also those of the task force," he says.

Töpfer says he has learnt a lot since becoming executive director of UNEP four months ago. But he adds that he will not be distracted from his reforms, just as his predecessors remained, despite much opposition, committed to setting up the IPCC.

"They didn't say 'let's change our minds' because of the criticism," Töpfer says. "Don't worry about the criticism. If someone is not criticized, it means he is not doing anything. We are doing things, which is why we will be criticized. But we also get applause."

The UN task force recommended setting up a new Environment Management Group comprising representatives of all relevant UN institutions and chaired by Töpfer. It also suggests reviving Earthwatch, the UN's system of gathering environmental information, set up in the 1970s.

Ehsan Masood

UNEP

French panel calls for closer monitoring of genetic modification

[PARIS] A French lay panel has delivered a mixed verdict on the use of genetically modified organisms (GMOs). It concluded that more research is needed before the risks can be properly assessed, and that greater public monitoring is required to ensure that government policy on GMOs is not unduly influenced by economic interests.

This was the outcome of France's first 'consensus conference' or *conférence de citoyens*, organized in Paris two weeks ago along the lines of the model invented in Scandinavia and since applied in the United Kingdom and elsewhere.

The 'jury' of 14 lay people, screened for independence from interested parties such as environmental pressure groups or food-based industrial organizations, were coached on the subject in advance. During the conference, GMO-related issues were debated in front of them by scientists, non-governmental organizations, industrialists and other bodies.

In its report, the jury acknowledged that GMOs have potential benefits in medicine, and supported greater research in this area.

But it also argued that research into the ecological risks of GMOs is inadequate and requires a substantial new impetus.

In particular, the panel recommended that such research should be carried out by national research agencies. "The power of public research bodies is probably the best guarantee of independence with respect to private sector research and the influence of multinationals," it said.

It called for an improvements in both the composition and the procedures of the Commission de Génie Moléculaire, the body that advises the government on approving GMOs. In the meantime, several members of the jury felt that a moratorium of the use of GMOs would be appropriate.

The dominance of multinational firms was identified as a major concern. The jury argued that this risked leaving farmers dependent on the decisions of a few large monopolies. But it thought GMOs might improve the international competitiveness of France's agricultural produce.

Commenting in the newspaper *Le Figaro* on the outcome of this first consensus

conference, Claude Allègre, the minister for national education, research and technology, said there was a need for "a happy medium between the competence of the aristocracy and the dictatorship of ignorance". But he argued that, although caution was warranted, too much analysis could lead to paralysis. "If we had held meetings in the factories to know whether or not this or that was risky, plastic would never have been invented," he said.

The conference is widely seen as part of an effort by the government to ensure greater public consultation on technological choices, since trust in expert committees has been shaken by the recent crises over BSE and contaminated blood.

GMOs are a hot potato for socialist prime minister Lionel Jospin's government, which has provisionally authorized the cultivation of transgenic maize. The Greens, France's main ecological party, form part of his government, but agriculture is one of the largest sectors in the French economy, and those who work in the countryside are an influential block of voters.

Eric Glover

Senators back bill to double R&D funds over 12 years

[WASHINGTON] Eight senators have introduced a revised proposal to double US investment in research and development during the next 12 years and introduce strict guidelines for evaluating research programmes and eliminating inefficient ones.

The proposed law, S2217, replaces a previous proposal, S1305, which would have doubled research spending in ten years, but was supported by only 19 of the 100 senators.

To attract broader support, the bill sets less ambitious funding targets and incorporates new policy principles and evaluation requirements (see *Nature* 393, 504; 1998).

It also sets one aggregated target for all non-defence funding agencies, allowing money to be redistributed between science agencies. S1305 was criticized for specifying equal increases for all agencies, regardless of their mission or performance.

The bill, introduced on 25 June by Senator Bill Frist (Republican, Tennessee), is supported by a range of engineering and scientific societies and the universities. But it has not won the active support of the biomedical research community, which is pursuing its own plans to double funding for the National Institutes of Health in just five years.

David Moore, a lobbyist at the American Association of Medical Colleges (AAMC), is uncertain whether the AAMC would even discuss endorsing the bill. "I think we'll continue pursuing the doubling of the biomedical research budget in five years," he says.

But Frist said he thought the bill would attract more sponsors than the previous proposal, and would be approved by the Senate. As an 'authorization bill', the measure sets non-binding targets for future spending.

Phil Gramm (Republican, Texas), who sponsored the previous proposal, said the new bill would now stand a better chance of progressing because Frist would carry it through the subcommittee that he chairs — the science, technology and space subcommittee of the Senate commerce committee.

Scientific societies hope that a companion bill will be proposed in the House of Representatives soon after Vernon Ehlers (Republican, Michigan) completes his review of science policy later this month. They hope that the leadership of the House, including Newt Gingrich (Republican, Georgia), who has recently made several speeches in support of increased research funding, will support such a bill.

Allan Bromley, former president of the American Physical Society and presidential science adviser, thinks the new bill will be passed by the Senate this year. Frist says its prospects depend on how many sponsors it attracts in the next two weeks. **Colin Macilwain**

NASA loses touch with solar observatory in space

[WASHINGTON] Ground controllers lost contact last week with the \$1 billion Solar and Heliospheric Observatory (SOHO), the centrepiece of an international programme to study the Sun and its interaction with Earth. Scientists involved in the US-European project were pessimistic at the beginning of this week about its chances of being recovered.

Engineers at the US space agency NASA's Goddard Space Flight Center were performing routine maintenance on 24 June of SOHO's position-keeping ability when the spacecraft lost its orientation to the Sun.

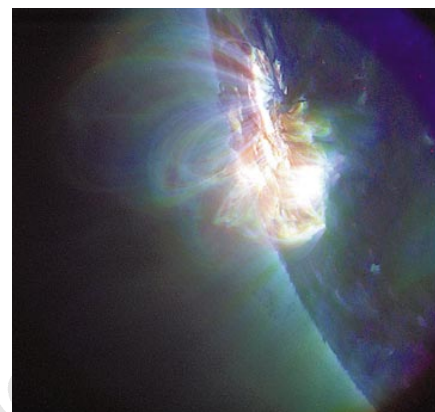
Although it went into an automatic routine designed to reorient itself to the Sun, the spacecraft lost radio contact with the Earth. Attempts to regain contact were unsuccessful, even using the large dish antennas of the Deep Space Network. If its solar arrays were pointing in the wrong direction at the time, its on-board batteries would have drained quickly, leaving it without power.

Launched in December 1995, SOHO had completed its nominal two-year mission, but the European Space Agency (ESA) and NASA, who shared the development costs, had extended its operations to 2003.

The decision had strong backing from the scientific community. A report from the National Research Council urged NASA to continue operating SOHO and the other satellites of the International Solar Terrestrial Program (ISTP), including WIND and POLAR, through the "solar maximum", the peak in the 11-year cycle of solar activity expected between 1999 and 2002.

SOHO has returned important results ranging from the detection of solar quakes caused by flares to a detailed study of coronal mass ejections. Parked in a stable orbit between the Sun and Earth, it monitors the Sun's surface and extended atmosphere.

Failure to recover the observatory would



Solar power: a false-colour image from TRACE, which may take on some of SOHO's tasks.

be a "tremendous loss" to the solar-maximum observing campaign, says project scientist Art Poland. No other satellite has spectrometers and coronagraphs trained on the Sun, or can return data on solar seismology and magnetic fields.

NASA's recently launched Transition Region and Coronal Explorer (TRACE) spacecraft could offset the loss using its ultraviolet telescope, according to Alan Title of the Stanford Lockheed Institute for Space Research in Palo Alto, California, a principal investigator for TRACE and member of the SOHO team. But it would have to record data more frequently, he says, because it has a narrower field of view than SOHO.

Plans would probably involve placing duplicate instruments on a variety of different spacecraft, rather than relying on another single observatory such as SOHO.

One investigator thinks the capability of the mission could be duplicated for 20–30 per cent of the cost of SOHO. But NASA and ESA would have to act quickly, and funding would have to be approved. **Tony Reichhardt**

Space agency to work closely with EU

[MUNICH] The council of the European Space Agency (ESA) last week approved a resolution agreeing to work more closely with the European Union in developing strategies in telecommunications, navigation and Earth observation.

The council also approved the launch of a programme for Earth observation, known as Living Planet. A total of ECU24 million was agreed by 10 member states plus the associated member state, Canada. Although it is 20 per cent less than the designers asked for, programme director David Southwood says it is sufficient "to move ahead immediately". Other member states may

join the programme later, he adds.

The Living Planet programme will organize a series of "small and frequent" missions in areas such as oceanography, climatology and atmospheric chemistry, to be launched after 2002. Opportunities for small missions in these categories will be announced in the next few weeks.

The programme was designed after discussion with the European Commission, reflecting the willingness of the two bodies to work together more closely. Southwood says the resolution "prevents us from going off in different directions, or from competing with each other". **Alison Abbott**

New doubts over rock-dating techniques

[SAN DIEGO] The rock-dating techniques developed by a geographer at Arizona State University (ASU) have again been called into question as a detailed analysis published last week strongly suggested that some of the results may have been fabricated.

The dating method devised and used by ASU's Ronald I. Dorn was shown to be largely meaningless in an article published in *Science* by researchers from four institutions, including the University of Arizona in Tucson.

In an accompanying rebuttal, Dorn acknowledged that his method for dating rock surfaces was fundamentally flawed. But he denied any impropriety, calling the suggestions of sample content manipulation “utterly false”.

After the publication of the *Science* article, a geologist from a fifth institution — John W. Bell of the University of Nevada at Reno — told *Nature* that he had discovered the same type of irregularities with Dorn's work as had been identified by the Arizona-led group.

Bell's discovery, involving rock varnish samples for a research project near the proposed Yucca Mountain nuclear waste site in Nevada, seem to provide independent confirmation of the findings of the Arizona-led group, raising further questions about the credibility of Dorn's research.

"We are worried about the scientific integrity of the work [Dorn did for us]" said Bell. "If there is any bad data, we want to make sure we weed that out." He said that excluding Dorn's data is not expected to change the results of his study, which examined the age of an earthquake near the waste site.

Dorn's method involves scraping varnish



off a rock surface, treating this sample, and then analysing the remaining substance using accelerator mass spectrometry (AMS) to determine its age. Working with scientists around the world, Dorn has dated surfaces near rock art in the United States, Australia and Portugal, and used the same method for an earlier published study on rock formation age near Yucca Mountain.

The article in *Science* revealed that an analysis by two AMS laboratories found pieces of coal and/or charcoal-like material in about 80 of Dorn's samples from previous projects. The coal and charcoal have widely disparate radiocarbon dates, say the scientists, who also are from Columbia University's Lamont Doherty Earth Observatory, Northern Arizona University and Eidgenössische Technische Hochschule in Switzerland.

“If you have a sample mixture of two different ages, it will not yield a reliable age, just a measure of a ratio for the age of the two com-

ponents,” says J. Warren Beck, a University of Arizona geochemist and a lead author of the *Science* article. Beck and colleagues found no coal or charcoal in rock varnish samples they collected and analysed to try to duplicate Dorn’s work.

After first making the unusual discovery in 1996, the University of Arizona scientists reported the findings to the National Science Foundation (NSF), which funds both their AMS laboratory and Dorn's research. Officials say that the NSF then subpoenaed Dorn's samples from the University of Arizona, and ASU began an inquiry into whether Dorn had committed scientific misconduct by altering the samples with coal and/or charcoal to fabricate dates.

But the probe dragged on until details of the dispute were published in *Nature* (392, 218; 1998) earlier this year. ASU officials in Phoenix then began interviewing Arizona scientists involved in the controversy. No administrators from ASU, Arizona or the NSF would discuss the current status of the Dorn inquiry.

In his rebuttal, Dorn insists that coal and charcoal occur naturally in rock varnish, writing “my results have been fully replicated by others”. In particular, Dorn cites an unpublished article by ASU geologist J. Ramon Arrowsmith, who Dorn writes “replicated in an independent study” his work after being “trained in sample collection and preparation procedures”.

But in an interview last week, Arrowsmith denied that his article replicates Dorn's rock-varnish method. Arrowsmith says he used a different technique to look for organic material in "the contents of cracks" in rocks, that no AMS dating tests were performed on the sample, and that he wasn't trained by Dorn to use his established method of dating rock varnish. Arrowsmith described his project, started last autumn, as "a quick way" of trying to "show these particles could be found" in rock.

Scientists at the University of Arizona and Nevada remain sceptical about Dorn's claim of finding the organic material in the rock varnish, because the coal and charcoal-like pieces identified were so large. The Arizona scientists say they found pieces as large as 1 mm across. Bell said that a microscopic examination of 5 of the 15 Nevada samples — which were analysed at the Rafter Radiocarbon Laboratory in New Zealand — also found pieces of this size. Noting that the rock varnish was only 0.2 mm thick in the Nevada research, Bell said: "I have not heard any reasonable explanation of how particles that size could be extracted from such thin varnish."

Bell said that Dorn has refused to let him retrieve the Nevada samples from New Zealand for more analysis. Dorn was unavailable for comment, reportedly on a field trip.

German institute 'correct' to fire technician

[MUNICH] An investigation into a case of scientific fraud at the Max Planck Institute for Plant Breeding in Cologne has endorsed the institute's decision to dismiss both a technician who admitted manipulating experiments and the leader of her research group, Richard Walden, who accepted overall responsibility (see *Nature* 393, 293; 1998).

The investigation committee presented its report on the affair, which continued for six years, at the annual meeting of the Max Planck Society (MPS) in Weimar last week. But the report also concludes that Jeff Schell, the director of the institute's department of plant genetics in which the experiments were carried out, did not neglect his duty of ensuring good scientific practice.

Furthermore, Schell's authorship of many of the papers in question could not be regarded as merely honorary, says the report, as he made a significant intellectual contribution to the design of experiments

and discussion of their results. New MPS rules frown on honorary authorship.

Crucial experiments in the papers under suspicion, published in high-profile journals including *Nature* and *Science*, were repeated during the four-month investigation. The report has been sent to the external chairman of the investigation committee, Walter Odersky, a former president of the German constitutional court, and Hildegund Holzheid, president of the Bavarian constitutional court, who have been asked to judge if any issues have been left open.

The MPS committee on good scientific practice is considering what lessons can be learnt from the case. "We want to encourage more openness between scientists of an institute, so that new results are widely discussed to the point where experiments could be repeated [for confirmation] by colleagues," says Hubert Markl, president of the MPS.

Alison Abbot

Rex Dalton

China plans major shake-up of academy

[BEIJING] China's leaders have approved a plan by the Chinese Academy of Sciences to carry out a drastic re-organization of its 123 institutes, which employ 68,000 researchers.

The plan will be implemented in three phases over the next 13 years. Substantial amounts of new money will be directed into selected research areas and regions, with hundreds of new staff being recruited from overseas. Some institutes will be merged, some closed down.

In line with government priorities, the academy aims to "embrace the era of knowledge economy" and "build up a national innovative system". To achieve this, it is seeking an extra 2 billion yuan (\$240 million) a year — increasing by half its institutes' current funding from a number of sources, such as the National Natural Science Foundation of China and the Ministry of Science and Technology, as well as the academy itself.

The proposed extra budget has yet to receive official approval from the financial authorities. But China's top committee for science, technology and education, headed by Premier Zhu Rongji, pledged its support

last month and the plan won backing earlier this year from President Jiang Zemin. A decision on the budget is expected in about a month's time, according to Bai Chunli, one of the academy vice-presidents.

In the first phase, from 1998 to 2000, extra resources will be invested in priority research areas and regions (see table). Research groups from about 40 of the academy's 123 institutes are expected to participate, and 600 new staff will be recruited from outside the academy.

It is hoped that many of these will be bright young researchers from overseas, says Bai. As well as basic scientists, the academy will be seeking individuals with good research management skills.

New research centres will be created in the selected fields and regions by combining strong groups from different institutes. It is hoped to create 10,000 positions (more than double the current number) for 'mobile' researchers such as postgraduate students, postdoctoral fellows and visiting professors.

The aim is to reduce the 68,000 permanent positions to 30,000 through retire-

Priority research areas and regions, first phase (1998-2000)	
Location	Research area(s)
Shanghai	life sciences
Beijing	material sciences
North-west (Lanzhou)	natural resources, environment, and sustainable development
South-west (Kunming)	biological resources, biodiversity, conservation
Shanghai	high technology
Beijing	information technology
North-east (Changchun)	high-performance materials, advanced manufacturing technology

ments and re-assigning people to other jobs by 2010, says Bai, while increasing the number of mobile researchers to 30,000.

A 'tenure track' system will be introduced for full-time employees, with assessments every four years up to the twelfth year, before a decision on lifetime employment is made. Details of the first phase of the re-organization plan, such as which institutes will participate in the priority areas, will be finalized in about two months.

David Swinbanks

Australia builds on biological successes with a broad new centre

[SYDNEY] A broad-based Biological Sciences Institute is being set up at the University of Queensland, Brisbane. Unusual in scope outside the pharmaceutical industry, it will be the best-funded new research facility established in Australia for many years.

A four-way financing deal, raising A\$55 million (US\$33 million), was sealed last month, three days before a closely fought State election. Australia's federal science minister, John Moore, whose electorate covers the campus, announced a grant of A\$15 million, matching one pledged a year ago by Rob Borbidge, then Premier of Queensland's coalition government.

The federal contribution is the first grant to be made for scientific purposes by the Commonwealth Federation Fund, which usually finances large-scale public works such as roads, railways and dams. The university is contributing A\$15 million and the university's vice-chancellor, John Hay, has announced a commitment of A\$10 million from an anonymous private source, thought to be from North America.

The institute will house divisions in genomic research, developmental biology, cellular biology, structural biology, molecular design and bioinformatics. A 14,000 square-metre building is due to be completed in two years. An initial staff of 350 is expected to double by 2005.

The Commonwealth Scientific and Industrial Research Organisation has



Designs on biology: John Mattick, seated, John Hay and acting manager Emma Puttick.

expressed interest in joining the project. Its own staff would focus on agricultural biotechnology — including plant and animal genomics — and may add up to 8,000 square metres more research space.

Project leader John Mattick describes the scheme as a "maxi form of Co-operative Research Centre", a reference to Australia's scheme combining research by universities, government agencies and industry.

Four existing centres at the university will be incorporated into the new institute. These are Mattick's Centre for Molecular and Cellular Biology, the Centre for Drug Design and Development headed by Peter Andrews, the Centre for Microscopy and Microanalysis and the Australian National Genome Research Facility. Alliances with

drug companies are expected, and the institute will provide substantial space for other research organizations.

The centres have already been involved in a range of projects, often with national and international collaborators. These have included bacterial, plant and mammalian genome sequencing. Their work in molecular genetics and in developmental, cancer and cell biology has already led to the identification of several developmentally important genes, and staff have investigated the dynamics of cell membranes (see *Nature*, 384, 427-432, 1996 and 392, 193-197, 1998).

The structural basis of novel venoms and toxins has been unravelled from cone shells and other tropical species. High throughput screening of products from Queensland's coral reefs and rainforests is giving promising leads in the development of novel drugs for central nervous system and cardiovascular disorders.

The timing of the announcement raised criticism that it was a last-minute effort to gain political credit before the State election. (In fact, Borbidge was defeated, in a result that rocked Australia's Liberal government.) But Moore has denied that political considerations were at work. Mattick says the project has always enjoyed bi-partisan support, and is confident that Queensland's new Labor Premier, Peter Beattie, will remain committed to the funding. Building work is due to start shortly.

Peter Pockley

Organic farmer takes gene battle to court

[LONDON] A dispute over the distance pollen can travel while remaining able to cross-pollinate effectively with other plants has become the latest twist in the British debate about genetically modified crops.

The issue is central to a court challenge being mounted by a farmer who is trying to stop an experimental crop of genetically modified maize being planted next to his farm in Devon. The farmer, Guy Watson, is being supported by the environmentalist group Friends of the Earth and by the Soil Association, Britain's largest certifier of organic produce.

The three have joined forces out of concern that the maize, which has been modified to be resistant to a herbicide, may pollinate a field of organic sweetcorn planted two kilometres away on Watson's farm. The Soil Association says it will withdraw Watson's 'organic' certificate if the trial goes ahead.

But Sharpes International Seeds Limited, the company responsible for the trials, argues that the farmer and his supporters are unduly concerned. Along with the National Institute of Agricultural Botany in Cambridge — which is conducting the trial on Sharpes' behalf — the company contends that the likelihood of pollen from the modified field contaminating the organic sweetcorn is too small to be measured.

This, they say, is because the distance between the two fields is 10 times the minimum legal requirement to achieve the 99.9 per cent seed purity that a crop needs for organic certification.

The court case is the latest hurdle to attempts to commercialize genetically modified crops in Britain, in the wake of mounting public concern about their safety.

Some local authorities are refusing to serve genetically modified food in school meals, and some supermarket chains will not stock unlabelled genetically modified products imported from the United States.

Conservation groups are advocating a complete moratorium on commercialization until important questions can be answered. Last month Prince Charles expressed deeply held beliefs against the genetic manipulation of food.

The resistance to genetic modification of crops is not just passive. Experimental crop sites are frequently vandalized, leading to producers calling on the government not to make their location public on the grounds that they are being prevented from carrying out research that will help to establish whether genetically modified crops are safe or not (see *Nature* 393, 726; 1998).

The government has found itself torn between those who want to push ahead with commercial-scale crops and those who want to tread a more cautious line. The Depart-



Hope or threat? Farmers fear for their organic certification from genetically-modified pollen.

ment of Trade and Industry is keen to support the biotechnology industry and wants the regulatory bureaucracy to be kept to a minimum. But the science minister, John Battle, appears to be toeing a cautious line and has plans to hold a 'consensus conference' on the issue later this year.

The Ministry of Agriculture is split between those who consider that genetic modification is integral to the future of agriculture, and those who want to resolve concerns for food safety before the technology becomes widespread.

The latter group includes members of the UK Register for Organic Food Standards (UKROFS), an agency appointed by agriculture ministers that decides whether a method or product can be called 'organic'.

The agency is based within the agriculture ministry and has so far maintained total opposition to genetically modified crops being labelled organic. It has not said whether it supports Watson's legal case, but it says that genetically modified organisms "have no place in organic production systems".

The Department of the Environment has to decide, on the basis of trial results, whether a crop should be commercialized. Commercialization also needs the consent of each European Union member state.

The decision on how to respond to Watson's challenge will be taken by Michael Meacher, the minister of state at the Department of the Environment. Meacher is sympathetic to environmental arguments, but his decision will be made more complicated by members of his scientific advisory committee on releases to the environment (ACRE).

Many committee members doubt if the plot of modified maize could pollinate Watson's organic sweetcorn crop, although they do not completely discount the idea. The committee's advice will not be made public

until the minister issues his formal response.

The views of ACRE members are well known to those following the debate, which is why Friends of the Earth, the Soil Association and Watson chose to go to court before Meacher's formal response. They also want the trial to be halted before the flowering season begins this month, which is when pollination will take place.

Pete Riley, a biotechnology campaigner for Friends of the Earth, says the evidence for cross-pollination is overwhelming. He says that pollen can maintain its fertilizing ability for up to 80 hours after flowering.

In addition, Riley says that recent laboratory research points to a higher mortality rate among lacewings that feed on genetically modified maize, apparently contradicting an earlier field study. He also questions the overall desirability of a crop that has been genetically modified to be resistant to herbicide when research has shown that such a crop can transfer herbicide resistance to weeds growing nearby.

Scientists generally agree that some pollen can be long-lived, and that it can also travel distances of up to a kilometre.

But one of the main scientific difficulties in this issue is that the probability of cross-pollination, and its effects on the environment and food safety, is too small to be studied effectively. This, says Mick Crawley of Imperial College, London, is one of the biggest obstacles to further research.

What is known, says one ACRE member, is that the "vast majority" of pollen grains lose their fertilizing ability within half an hour, and do not travel distances much beyond 200 metres. Experience, he says, shows that a distance of 200 metres between two crops results in 99.9 per cent seed purity. "This limit is an internationally agreed standard based on practical experience. It has not been set by people sitting in an office."

For cross-pollination to occur, he adds, the two crops must flower at the same time, which he believes is unlikely. The sweetcorn must also be located downwind of the maize.

The Devon site contains six plots of modified maize, with some 1,200 plots of unmodified crops. The ACRE member says that the likelihood of cross-pollination from these unmodified sources is more likely than from one of the genetically modified sites.

On the issue of herbicide tolerance spreading to other plants, the ACRE member says that such a trait tends to appear "within a few generations" of a single crop which is constantly sprayed with herbicide.

"The trouble with the opposition to genetic modification is that it is rooted in too much Plato, and not enough Aristotle", he says. "There is not enough of the detail, and too much of the general."

Ehsan Masood

Corporate investors take US research bonanza to \$200bn

[WASHINGTON] Total US spending on research and development surged to more than \$200 billion for the first time last year, fuelled by stronger investment in corporate research and development (R&D).

According to *Science and Engineering Indicators 1998*, the latest biennial guide to research statistics compiled by the National Science Foundation, the total grew by 4 per cent more than the rate of inflation last year.

The new statistics confirm a trend that began around 1994 when, after a decade of virtual stagnation, total US R&D spending began to surge forward, driven by private sector investment despite a virtual standstill in funding from the federal government.

Since then, the total US investment in non-defence R&D as a percentage of gross domestic product — perhaps the most reliable measure of national R&D performance — has risen steadily from 1.95 per cent to 2.15 per cent.

Spending on basic scientific research reached an all-time high of \$31.2 billion last year, up by 3 per cent more than inflation. Some \$18 billion of that came from the federal government, but the industrially-funded component rose most sharply, to almost \$8 billion.

Max Planck creates fund to lure top scientists

[MUNICH] Germany's Max Planck Society (MPS) is selling off property worth DM10 million (US\$5.5 million) to create a fund to finance fringe benefits for top scientists it wants to attract to its institutes. The plan was drawn up earlier this year to coincide with the society's fiftieth anniversary (see *Nature* 392, 3; 1998). The details were announced at last week's annual meeting of the society in Weimar.

The fund, called *Exzellenzsicherungsfonds*, is intended to restore the competitive edge of the MPS in attracting leading researchers to head its institutes. MPS is now trying to increase the value of the fund to DM20 million. The steel company Krupp has already announced a donation of DM1 million.

French minister sets up science advice panel

[PARIS] France is to establish a 'national council for science' to advise the government on scientific and technological issues, such as genetically modified organisms (see page 4), spongiform encephalopathies, the management of nuclear waste, and the hole in the ozone layer.

The move was announced by Claude Allègre, the minister for national education, research and technology, in an interview this week with the newspaper *Le Figaro*, organized in the run-up to an interministerial meeting on research scheduled for 9 July. Allègre indicated that he would try to give the new body responsibility for fixing research strategy, a task presently carried out mainly by research organizations such as the Centre National de la Recherche Scientifique.

Paris's 'genetics valley' wins funding boost

[PARIS] Efforts to create a 'genetics valley', bringing together public laboratories and private companies at Evry outside Paris, have received a boost with the approval by the region of Essonne of FF30 million (US\$5 million) for the project, almost half of which will be used to build a conference centre.

The site, known as Génopole, is centred around the Généthon laboratory which made large strides in early genome mapping efforts (see *Nature* 372, 397; 1994), and is home to the national sequencing and genotyping centres. Efforts to attract others to the area are being coordinated by Pierre Tambourin, a former director of life sciences at the Centre National de la Recherche Scientifique.

UK universities set to bid for venture capital

[LONDON] The British government this week released details of its £50 million (US\$80 million) University Challenge Fund designed to help promising research projects emerge from the laboratory to the marketplace. A competition will be held later this year in which universities will bid for seed venture capital funds. Each scheme will need to be led by a fund manager.

The fund was announced by Gordon Brown, the Chancellor of the Exchequer, in his budget in March. Money has been committed to the fund by the government, the Wellcome Trust, and the Gatsby Charitable Foundation, one of the charitable trusts of the Sainsbury family.

Africa seeks to assert rights to 'natural' drugs

[LONDON] The governing council of the Organization of African Unity (OAU) has approved a controversial model bill for its member states on the issue of bioprospecting. The bill refuses to recognize any patent on a drug made from natural products found in Africa — unless it acknowledges the 'ownership' and contribution of the relevant community to the new product.

The OAU council has also called for the

adoption of an 'African Convention on Biological Diversity' to safeguard the interests of local communities. The model bill was drafted by Ethiopia with the OAU's scientific and technical research commission (see *Nature* 392, 423; 1998). It states that ownership of new compounds should rest with indigenous local communities for "all time and in perpetuity".

Physicists back study of UFO evidence

[WASHINGTON] Believers in exotic phenomena received encouragement this week when a panel of scientists set up by a professor of applied physics at Stanford University issued a report arguing that it might be "valuable" to evaluate carefully reports of unidentified flying objects (UFOs). This conclusion is being contrasted to that of an earlier panel, which reported 30 years ago that further extensive studies of UFOs "probably cannot be justified".

The new report, published by the Society for Scientific Exploration, was based on evidence for the existence of UFOs presented by eight investigators. But the panel, organized by physicist Peter Sturrock, warns that evaluations of the evidence "must take place with a spirit of objectivity and a willingness to evaluate rival hypotheses" that has so far been lacking.

Royal Institution names its first woman director



[LONDON] Britain's Royal Institution has appointed Susan Greenfield (above), professor of pharmacology at the University of Oxford, as its first woman director. In 1994, she became the first woman to deliver the institution's famous annual Christmas lectures. She will take up her post on 1 October, succeeding the chemist Peter Day.

Greenfield is a prolific popularizer of science and a familiar figure on radio and television, who is also active in research. She heads a group looking into diagnosis and therapy for neurodegenerative disorders. She also writes a weekly newspaper column, and is due to present a six-part television series on the brain and mind.

Let's share the excitement of science

Sir — Your editorial “Dangers of publication by press conference” represents an excellent opportunity for open discussion of the peer-review process and how best to present scientific results to the public (*Nature* 393, 397; 1998). We also appreciate the compliment that the US space agency NASA has played an important role in increasing public interest in science.

You argued that public release of science results should occur only after peer review by journals. Without it, journalists may be unable to do justice to the story, and perhaps “the public will more often be faced with sensational and triumphant stories that subsequently prove to have been false”. You concluded that NASA risks “undermining the respect for objectivity on which the public support for science ultimately rests”. This narrow conclusion gives short shrift to public interest and the skills of the media.

NASA's charter requires the agency to disseminate as widely as possible the results of its publicly funded research. In the process, we must and do scrupulously safeguard our scientific credibility through strong attention to the peer-review process. Most of NASA's press briefings and press releases feature peer-reviewed results.

However, the benefits of having a larger

vision are evident in many recent examples of near real-time science being presented to the people who've paid for it. Consider public awareness and excitement about science spurred by announcements of possible subsurface oceans on Europa and ice on the Moon, a glorious image of the Eagle Nebula, and breaking news such as comet Shoemaker–Levy 9 and Mars Pathfinder. There is little evidence of damage to public support for science in the wake of these often tentative findings — indeed, quite the opposite.

When a NASA-funded researcher chooses to submit a paper at a scientific meeting before journal publication — as with Susan Terebey's possible extrasolar planet — we consider its merits on a case-by-case basis. Usually, the results are scrutinized intently by scientists independent of NASA before we decide to release the results, which *Nature* acknowledged constitutes a form of peer review. My office turns down four or five interesting scientific results for every one that we decide to present publicly.

In subsequent media activities, we strongly underscore the preliminary status of the results, and the need for further observations. We work hard to ensure the

story is told correctly. This message was clearly received during the press briefing on Terebey's research, as every one of the dozens of press reports that we saw included the proper caveats. Journalists were indeed able to do justice to the story and, in addition, probably did more to educate the public about the scientific process than any amount of journal editorializing ever could.

The tax-paying public is entitled to receive scientific results expeditiously in simple, understandable language. Non-scientists enjoy participating in the excitement and adventure of science — it may be preliminary and uncertain, but letting them participate in the progress of science can only be beneficial.

Finally, if the so-called “blob” in the image Terebey obtained is later proven not to be a planet, NASA will endeavour to be the first to announce it. We trust the public's common sense and its ability to understand that progress is not made without some missteps. Based on the NASA values of openness and honesty, we are confident in the correctness of our philosophy.

Edward J. Weiler (Scientific Director)

The Astronomical Search for Origins and Planetary Systems, Code S, NASA Headquarters, Washington DC 20546, USA

Biotech battlelines

Sir — When the Swiss people voted in the referendum of 7 June against the proposal to impose restrictions on genetic engineering, they signalled that they had decided against economic and scientific isolation from the rest of the world. But how was it possible that the fate of the rational world of science was decided by a war of emotions?

At the beginning of the referendum campaign, proponents and opponents of genetic engineering alike seized powerful images of human and animal suffering to rally the public behind their banners. Images of hospital patients were used by biotechnology proponents to remind the public that a ban would inflict a heavy price on those suffering from severe diseases and dependent upon state-of-the-art medical treatments. In 1996, 10 years after the Chernobyl explosion, those in favour of the ban circulated a portrait of a child born without eyes, with the words: “No one can predict the damages (from technology), as was the case for Chernobyl. We do not want a genetic catastrophe!”

As the campaign progressed, there was a shift in tactics on both sides. Those against the ban realized that complex scientific and moral considerations could be

communicated to the public by emphasizing the importance of biotechnology to medical research, diagnostics and treatment. Proponents of the ban abandoned striking visual images, accentuating instead broader issues such as the dignity of creation and the purported ecological impacts of genetically modified crops.

The campaigning appears to have obeyed the principle revealed by the Eurobarometer survey of attitudes to biotechnology (*Nature* 387, 845–847; 1997). While people were willing to accept some risk as long as there was a perception of usefulness, the presence of moral doubts acted as a powerful veto by overwhelming all other considerations, including usefulness and risk.

The Eurobarometer study also demonstrated that, of the biotechnology applications it surveyed, genetic screening and medicines obtained the highest ratings for usefulness and moral acceptability, with a positive evaluation of risk. Crop plants were also evaluated positively for usefulness and moral acceptability, but the balance of risk was seen as negative. The authors of the study concluded that the technically based reassurances of regulators are unlikely to alleviate public concern about the acceptability of certain applications of biotechnology. The moral and political

dimensions of risk are missing from the objective risk assessment framework that is usually applied to making cogent decisions about new technologies.

What we have seen in Switzerland is that the rational and analytical approach of science is no longer consistent with other world views. The era of objective, value-free science in an ivory tower is coming to a close. Scientists can no longer afford to ignore the social framework within which their research is performed. During the referendum campaign, scientists were concerned about the possibility of losing an important research tool — transgenic animals. It was the seriousness of this threat that drew them into the public spotlight to explain the importance of their work.

By choosing not to isolate themselves even further, the Swiss have recognized that science, when guided by evidence, can transcend linguistic and cultural barriers. This aspect of science serves the egalitarian ideals of modern, pluralistic societies in which all citizens can expect to exercise their right to vote.

Othmar Kappeli & Lillian Auberson

Biosafety Research and Assessment of Technology Impacts, Swiss Priority Programme, Biotechnology, Swiss National Science Foundation, Clarastrasse 13, CH-4058 Basel, Switzerland

A conduit to the core

Richard W. Carlson

A better understanding of the dynamics of Earth's mantle depends on improved techniques for isotopic tracing. One such study has resolved minute variations in isotopes of osmium and links surface volcanism to events at the core.

For those fortunate enough to visit Hawaii, the thought of sitting on a volcanic conduit that extends 2,900 km into the Earth, to the boundary between mantle and core, may only briefly disrupt consideration of more weighty issues such as when to order the next Mai-Tai. Nevertheless, from data reported in *Science* by Brandon *et al.*¹, it seems that Hawaiian volcanic rocks may contain a small component of material derived directly from the Earth's metallic core. The discovery could be the long-sought evidence that the ultimate cause of mid-plate 'hotspot' volcanoes like Hawaii is narrow plumes of hot mantle rising from the boundary between core and mantle². This finding would link surface volcanism to core processes³ in the ultimate goal also served by the Mai-Tai, to cool the interior of a hot body.

When a fluid body is heated primarily from below, a strong temperature gradient develops at the boundary and causes hot, narrow plumes to rise upwards from the boundary layer⁴. This type of thermal boundary must exist between core and mantle, where the relatively low-density silicate mantle floats above the molten iron metal of the outer core; but if the mantle is layered compositionally, there might also be other such boundaries within the mantle. So identifying the depth of origin of hotspot volcanoes is essential to understanding the depth to which mantle convection extends, and the degree of chemical layering in the mantle. But this is no easy task. As described last year,

high-resolution seismic imaging of the mantle has finally revealed plume conduits beneath Iceland⁵, but the depth resolution of such experiments is not sufficient to trace the conduits from the surface to the core–mantle boundary.

Enter the isotopic tracer technique employed by Brandon *et al.*¹. Hotspot volcanoes have long been known to erupt lavas that are compositionally distinct from those erupted at ocean ridges. Much of the compositional difference between these lavas is believed to reflect the contribution to hotspot magmas of 'recycled' crustal materials (that is, oceanic crust and sediments that have been injected into the deep mantle by the subduction of tectonic plates at convergent plate boundaries⁶). This conclusion derives largely from the analysis of certain radiogenic isotope systems, for example samarium and its decay product neodymium, that are particularly sensitive to the presence of crustal material in the mantle because average crust has much higher abundances of these elements than does the mantle (Fig. 1). In contrast, the abundance of these elements in the core is quite low (Fig. 1), making most previous studies insensitive to detecting core material that might have been added to the hotspot source.

Brandon *et al.* examined the isotopic composition of osmium, an element that receives contributions from the radioactive decay of both rhenium (^{187}Re decays to ^{187}Os with a half-life of 42 billion years) and platinum (^{190}Pt decays to ^{186}Os ; half-life, 450

billion years). All three of these elements strongly concentrate into the iron metal of the core rather than the silicates that make up the crust and mantle (Fig. 1). Consequently, even small additions of core material to the mantle source of hotspots would dominate the isotopic composition of osmium, and should remain detectable even in the lavas produced by melting of the mantle plume when it nears the Earth's surface. Based on meteorite analogues⁷, the expected signature of core material in a hotspot magma is higher levels of ^{187}Os and ^{186}Os .

That hotspot lavas have higher relative abundances of ^{187}Os than does the upper mantle has been known for some time⁸. However, rhenium also is concentrated in the crust compared with the mantle (Fig. 1). So the relatively high abundance of ^{187}Os in hotspot magmas was initially attributed to the contribution of recycled crustal material⁸, supporting the model developed many years earlier⁶. The first studies of the osmium isotopic composition of hotspot lavas were not precise enough to resolve variations in the relative abundances of ^{186}Os , which are about 100 times smaller than variations in ^{187}Os .

Brandon *et al.* struggled for sufficient precision in their measurements and have managed to show that the ^{187}Os excesses in hotspot lavas are accompanied by correlated excesses in ^{186}Os . Platinum, however, is not enriched in crustal materials to the same extent as is rhenium (Fig. 1). So the observed co-variation of ^{186}Os and ^{187}Os in hotspot lavas is unlikely to result from mixing of subducted crust into the mantle, but is exactly what would be expected from mantle to which 0.5 to 1% core material has been added.

There is, however, a problem, in that this conclusion is difficult to reconcile with the lead isotopic composition of hotspot lavas⁹. Like osmium, lead is strongly concentrated in the core; but unlike osmium it is also concentrated in the crust relative to the mantle (Fig. 1). Uranium and thorium are the parents of the radiogenic lead isotopes. Because their concentrations in the core are expected to be very low, those of the radiogenic lead isotopes in the core should likewise be low. Hotspot lavas, however, do not show a consistent shift to unradiogenic lead isotopic compositions compared with ocean-ridge lavas, but rather show a tendency towards crustal compositions. Mass-balance calculations based on lead isotopic variations suggest that core input to hotspot sources must be very much less than 0.3% (ref. 9), compared with the osmium-derived figure of 0.5–1%.

Reconciling the evidence from osmium and lead remains a task for the future. Compared with osmium, estimates of the lead concentration in the core are quite uncertain, and there is only very limited information on

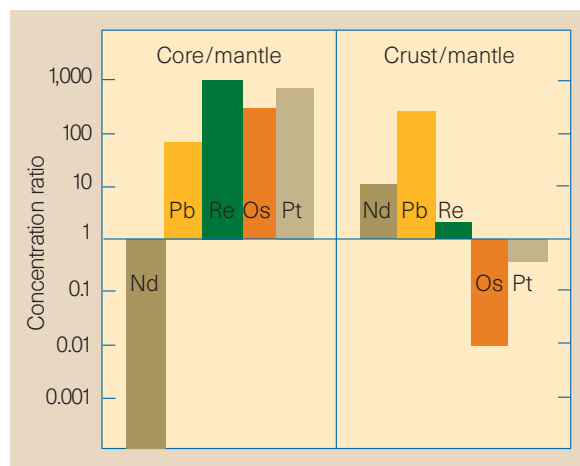


Figure 1 Concentration ratios, core/mantle and crust/mantle, for the elements neodymium, lead, rhenium, osmium and platinum. All of these elements participate in radiogenic isotope systems that are used to trace the interaction between chemically distinct materials in the Earth. In the work discussed here, Brandon *et al.*¹ have used osmium isotopes to infer that plumes of material rising through Earth's mantle to create 'hotspots' at the surface can originate as deep as the core–mantle boundary.

the distribution of rhenium, platinum and osmium in various crustal materials. Only recently have we had analytical techniques that are sensitive enough to examine the concentration variation of elements such as osmium at natural levels. All in all, then, these are early days. But the early results, such as those just published by Brandon *et al.*¹, should presage many exciting revelations about the nature of convection in the Earth's interior and the possibility that hotspot volcanoes reflect the direct transfer of heat from core to surface. □

Richard W. Carlson is in the Department of

Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington DC 20015, USA.

e-mail: carlson@dtm.ciw.edu

- Brandon, A. D., Walker, R. J., Morgan, J. W., Norman, M. D. & Prichard, H. M. *Science* **280**, 1570–1573 (1998).
- Morgan, W. J. *The Sea* **7**, 443–487 (1981).
- Larson, R. L. *Geology* **19**, 547–550 (1991).
- Sleep, N. H. *Annu. Rev. Earth Planet. Sci.* **20**, 19–43 (1992).
- Wolfe, C. J., Bjarnason, I. Th., VanDecar, J. C. & Solomon, S. C. *Nature* **385**, 245–247 (1997).
- Hofmann, A. W. & White, W. M. *Earth Planet. Sci. Lett.* **90**, 421–436 (1982).
- Walker, R. J., Morgan, J. W. & Horan, M. F. *Science* **269**, 819–822 (1995).
- Hauri, E. H. & Hart, S. R. *Earth Planet. Sci. Lett.* **114**, 353–371 (1993).
- Widom, E. & Shirey, S. B. *Earth Planet. Sci. Lett.* **142**, 451–465 (1996).

Vertebrate palaeontology

Evolutionary cut and paste

Neil Shubin

Fossil discoveries can muddle our attempts to construct simple evolutionary trees — fossils from key periods are often not intermediates, but rather hodgepodes of defining features of many different groups. The latest example of this phenomenon is announced on page 66 of this issue¹, where Jennifer Clack describes a 334-million-year-old fossil amphibian that displays, in the same animal, characters that are usually associated with three different types of primitive tetrapod. This discovery has implications for our understanding of how evolution proceeds and of how new groups of animals are assembled over geological time.

Extant amphibians and reptiles are distinct — anyone can tell a frog from a lizard. But add some fossil representatives of these groups, and much of their distinctiveness breaks down. This situation, where a fossil discovery reduces phylogenetic resolution rather than increases it, has recently arisen in several other contexts in vertebrate evolutionary biology. For example, living tetrapods have a variety of innovations that

separate them from fish. Perhaps the most famous of these are the loss of internal gills and the origin of digits, both of which are features usually associated with the invasion of land. Discoveries of internal gills in the tetrapod *Acanthostega*², and finger-like structures in the Devonian fish *Sauripterus*³, however, show the futility of relying on such characters to identify major groups. In both cases, the surprising implication is that these features may have arisen several times during the evolution of tetrapods. An identical problem is encountered in the evolution of birds and feathers. Although living birds are unique in their possession of feathers, recently discovered theropod dinosaurs show that different types of feather-like structures evolved several times in closely related theropods⁴. The main conclusion to draw from these discoveries is that 'key' characters can arise independently in different taxa and in different functional settings.

Clack describes a 15-cm-long four-legged creature that contains features that have previously been used to distinguish the

three major groups of early tetrapod: living amphibians and their relatives (temnospondyls), amniotes and their close relatives (anthracosaurs), and an enigmatic group known as baphetids (formerly known as loxommatids) which are vaguely crocodile-like in body plan and are characterized by having a unique keyhole-shaped orbit. The skull roof of this creature looks like that of a temnospondyl, the palate is like that of an anthracosaur, and the orbit is intermediate between that of baphetids and that of other tetrapods. Size-related variation adds some more mosaicism into the mix — small specimens have the primitively rounded orbits of other tetrapods, whereas those of the larger specimens approach the keyhole shape of baphetids. Clack performed a phylogenetic analysis of the new form and other Devonian and Carboniferous vertebrates, and its mosaic of characters leads to at least three conflicting phylogenetic arrangements (Fig. 1).

The mix of characters seen in Clack's creature reveals that the ecological experimentation during the early evolution of terrestrial vertebrates involved the independent evolution of similar features in closely related groups. A window on the palaeoecological context of this parallel evolution may ultimately be provided by the site where the fossil was found. Clack's is the latest in a string of remarkable discoveries to come from the East Kirkton Limestone (Lower Carboniferous). This limestone, exposed near Glasgow and Edinburgh, is one of the earliest definitively terrestrial localities⁵; it represents a shallow lake, and has produced a uniquely diverse fauna containing anthracosaurs, temnospondyls, scorpions and myriapods. This single site reveals that at least three different lineages of early tetrapod independently evolved into medium-sized fish-eating animals.

Parallel evolution may be a fact of life for palaeontologists, and a necessary consequence of how new structures evolve. Generally, it seems that major groups are not assembled in a simple linear or progressive manner — new features are often 'cut and pasted' on different groups at different times. Molecular studies point to a genetic mechanism for this pattern of parallel evolution⁶, in that the mosaicism seen in the fossil record may reflect the fact that similar regulatory genes are used in the developmental patterning of diverse types of animal⁷.

Molecular and developmental biologists are regularly faced with mosaics: similar genes and developmental processes are used to construct widely different types of organ and organism. Indeed, genetic evidence for parallel evolution comes from a comparative study of crustacean appendages⁸, in which changes in the pattern and organization of anterior appendages were found to correlate

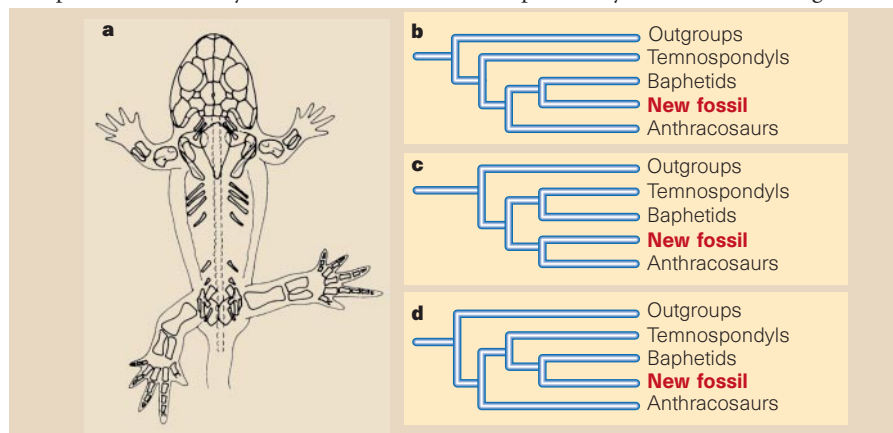


Figure 1 Palaeontological mix-up. a, The skeleton of the new amphibian described by Clack¹ has features that have previously been used to distinguish three major groups of early tetrapod. b, c, d, This mosaic of characters leads to at least three different phylogenetic interpretations.

with changes in the function of Hox genes. One implication of these results is that the independent evolution of similar types of appendage in crustaceans may have involved similar changes in regulatory gene function. This poses a 'chicken and egg' problem for palaeontologists: if independent evolution of key characters is common, how is phylogeny to be reconstructed? A way out of this quandary will probably only come from understanding the biological basis of the features that we use to reconstruct evolutionary

history.

Neil Shubin is in the Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6018, USA.

e-mail: nshubin@sas.upenn.edu

1. Clack, J. A. *Nature* **394**, 66–69 (1998).
2. Coates, M. I. *Trans. R. Soc. Edinb. Earth Sci.* **87**, 363–421 (1996).
3. Daeschler, E. B. & Shubin, N. H. *Nature* **391**, 133 (1998).
4. Ji, Q., Currie, P. J., Norell, M. A. & Ji, S.-A. *Nature* **393**, 753–761 (1998).
5. Clarkson, E. N. K., Milner, A. R. & Coates, M. I. *Trans. R. Soc. B* **87**, 417–426 (1994).
6. Carroll, S. B. *Nature* **376**, 479–485 (1995).
7. Panganiban, G. et al. *Proc. Natl Acad. Sci. USA* **94**, 5162–5166 (1997).

Particle physics

The Standard Model transcended

Frank Wilczek

Prolonged applause met Professor Kajita's presentation last month* of results from the SuperKamiokande detector¹. After decades of ardent searching, marked by numerous false alarms and several tantalizing but precarious hints, a physical phenomenon lying beyond the framework of the Standard Model of physics had been clearly identified. Neutrinos have mass.

The new observations concern the different types of neutrino (electron, muon and tau) produced by cosmic rays in the Earth's atmosphere, and they strengthen earlier indications² that there are fewer muon-type neutrinos than expected. In the details of its dependence on energy and angle, this anomaly is consistent with a very specific hypothesis: muon neutrinos are a mixture of two states with different masses. As time progresses, the theory goes, these two states fall out of phase; then the proportions of the mixture have changed, and no longer fit the specifications for a muon neutrino. The new mixture is part muon neutrino, part something else (probably a tau neutrino) — the muon neutrino has 'oscillated'.

The extent of the oscillation should depend in a very specific way on the difference between the masses, the original proportions of the mixture, the neutrino's energy and the distance it has travelled. The distance is related to the neutrino's arrival angle, as it originated in the atmosphere; and the energy is related to the energy released in the detector. The predicted dependence of the oscillation on angle and energy matches the observed anomalies quite well, so it seems appropriate — to the extent that it is ever possible — to claim that the observations 'rule in' the hypothesis. The neutrino mass difference needed to explain the observations is minuscule: a few hundredths of an electron volt, or less than 10^{-7} times the mass of the electron.

But the raw phenomena do not begin to convey the discovery's significance. It cannot

be considered as a bizarre, tiny correction to our previous understanding; rather, it is a first step towards a more comprehensive, and much more beautiful, formulation of the fundamental laws of physics. For, as I shall explain, this neutrino mass confirms bold theoretical ideas about symmetry and unification of forces that arise from the deep structure of the Standard Model.

In the Standard Model, the quarks and leptons that are the building blocks of matter fall into five separate classes, or multiplets (Fig. 1). There are 'strong' interactions, mediated by colour gluons, which can transform particles within a multiplet; 'weak' interactions, mediated by W bosons, that can transform particles between multiplets; and 'hypercharge' interactions, mediated by a mixture of photon and Z boson, which leave the particles unchanged (and include the more familiar electromagnetic interaction). It would be difficult to overstate the power and practical success of the Standard Model, but one cannot deny the superficially graceless cast of Fig. 1. Because we are discussing the most basic laws of nature, we have a right to expect better.

Fortunately, postulating a higher degree of symmetry produces a prettier picture^{3,4}. One version, called SO(10) symmetry⁵, looks especially good in light of the new neutrino mass measurements. In this unified theory (see Box, overleaf), all the particles of the Standard Model appear in one multiplet, and transformations within it include both the strong and weak interactions, now appearing on an equal footing.

The SO(10) unification of matter into one multiplet is a stunning achievement, but there have always been two nagging questions about it. First, one requires an additional particle to complete the symmetry, here called N. Where is it? Second, the postulated symmetry between the strong and weak interaction requires that they have equal power, which is not so.

The predicted properties of N are peculiar. Considering only the interactions of the

Standard Model (instead of the full SO(10)), N is completely neutral — it has no strong, weak or electromagnetic interactions. That makes it elusive. Related to this neutrality is a unique feature of N's mass. Whereas all the other particles would be massless in empty space, and acquire mass only as a result of their interactions with the omnipresent so-called Higgs field, N has an independent mass. Indeed, the interaction of N with the Higgs field converts it into an ordinary neutrino. All potentially observable consequences of the existence of N depend on the value of its mass. Fortunately, insight into this arises from our second nagging question — the inequality between strong and weak forces.

The coupling strength of forces in the Standard Model depends on distance in a precisely calculable way^{6,7}. By extrapolating these calculations to much smaller distances — equivalently, much larger energies — we can test the idea that the strong, weak and electromagnetic interactions are different facets of one universal interaction⁸. Remarkably, it works. The various coupling strengths do meet, indicating that unification occurs at distances below 10^{-32} m, or energies above 10^{16} GeV. In comparison, direct measurements currently peter out at around 10^{-18} m, or 100 GeV.

Now I can show how the different strands knit together to produce neutrino masses^{9,10}. N's mass is closely related to the unification scale — a very large mass scale by other standards. Ordinary neutrinos then acquire mass indirectly through their interaction with N: it is possible for a left-handed neutrino (ν) to change into a left-handed N, then a right-handed anti-N, and finally into a right-handed anti- ν (see Box). This process requires a quantum fluctuation to account for the fleeting existence of N and anti-N. But quantum fluctuations to states of such large energy are very rare. So the process whereby ordinary neutrinos flip their handedness, which is a measure of their mass, is heavily suppressed. Quantitatively, this argument

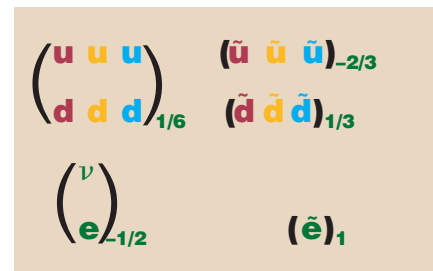


Figure 1 Multiplets of the Standard Model of particle physics. The up and down quarks, neutrino and electron shown here form only the first particle family (of three). The figure also omits right-handed particles, but the same pattern of five isolated multiplets is repeated for each family and handedness. The strong nuclear interaction transforms particles horizontally; the weak interaction, vertically. The subscript is hypercharge.

* Neutrino '98, Takayama, Japan, 4–9 June 1998.

Developmental biology

Boning up on Hedgehog's movements

Philip W. Ingham

Members of the Hedgehog (Hh) family of secreted proteins are known to act as intercellular signals in a multitude of processes, ranging from neuronal specification to bone morphogenesis¹. In the fruitfly *Drosophila melanogaster*, six genes have so far been implicated in transduction of the Hh signal, and the human homologues of three of these have been identified as oncogenes or tumour suppressors in the skin and nervous system¹. On page 85 of this issue, Bellaïche *et al.*² report that they have discovered a seventh gene that is involved in Hh signalling in *Drosophila*. The authors call this gene *tout-velu* (*ttv*) — a reference, in French, to the mutants' 'hairy' appearance — and show that it mediates movement of the Hh protein across responding cells in the *Drosophila* wing. A twist to the story is that *ttv* is homologous to *EXT-1* and *EXT-2* — two genes that are associated with benign bone tumours (exostoses) in humans^{3,4}. These findings raise intriguing questions about the mechanism of Hh signalling, while shedding new light on the pathogenesis of human multiple exostoses syndrome.

A major puzzle of Hh signalling is how the protein exerts its effects over relatively

long distances — the chemical properties of Hh seem to favour close association with the surface of cells from which it is secreted. This association is mediated by a lipophilic modification that occurs in vertebrates during processing of the protein⁵ through covalent coupling of cholesterol to the mature signalling form of Hh (called HhN). Diffusion of HhN is also constrained because it is sequestered in responding cells by its putative receptor⁶, a multi-pass transmembrane protein known as Patched (Ptc). In the absence of Ptc, however, HhN has been shown to travel over many cell diameters, notwithstanding its lipophilic attachment. What Bellaïche and colleagues² have now discovered is that this ability of HhN to diffuse depends on another putative transmembrane protein (in this case, a single-span type II class) encoded by *ttv* (Fig. 1).

The *ttv* gene was originally identified through its *hh*-like embryonic phenotype⁷, so, in principle, it could have acted at any point in the Hh signalling pathway. Crucially, however, Bellaïche *et al.* discovered that the activation of Hh-target genes is not abolished in the absence of *ttv*; rather, such activation becomes restricted to cells im-

mediately adjacent to those secreting the protein. This gave a hint that it may be diffusion of Hh, rather than Hh signalling *per se*, that is disrupted in *ttv* mutants. To explore this possibility, the authors made genetically mosaic animals in which large clones of cells lack expression of both *ttv* and *ptc*. When such cells abutted Hh-secreting cells, the result was striking — instead of passing freely across the cells, as it would in the absence of Ptc alone, the Hh protein travelled no further than the first cell that it encountered.

In *Drosophila* epithelial cells, the Hh protein is localized basolaterally, accumulating in distinct, punctate structures that seem to be present in both secreting and responding cells^{8,9}. These characteristic accumulations are still associated with responding cells in the absence of Ptc, suggesting that the uptake of Hh into these cells (endocytosis) does not depend on its interaction with Ptc^{8,9}. So, could Ttv mediate endocytosis of Hh?

On the face of it, the identification of Ttv as a type II transmembrane protein fits well with such a function — binding of Hh to Ttv's exposed carboxy-terminal domain could facilitate internalization of Hh. But a study just published by McCormick *et al.*¹⁰ in *Nature Genetics* indicates that the connection between the two proteins may be less direct. These authors have discovered that EXT-1 — the closest human homologue of Ttv — is found in the endoplasmic reticulum, where it helps to regulate the synthesis and display of cell-surface heparin sulphate glycosaminoglycans (GAGs). Because GAGs have been implicated in receiving signalling molecules, it is plausible to speculate that Ttv is involved in the synthesis of a GAG that specifically interacts with Hh at the cell surface. Such an interaction might result in the endocytosis of Hh or, alternatively, it may facilitate the movement of Hh from one cell to the next by translocating it around the surface of the cell.

The *EXT-1* gene was originally identified through its association with hereditary multiple exostoses. Patients have short stature and many benign tumours, which are mainly found near the joints of the long bones¹¹. So what is the connection between diffusion of Hh in fruitfly wings and human multiple exostoses? During normal bone development a member of the Hh family, Indian Hedgehog (Ihh), is expressed in a region of the long-bone growth plate, where chondrocytes (the progenitors of cartilage) exit the cell cycle and begin to mature. The chondrocytes are thought to be maintained in a proliferative state by Ihh, through the Ihh-stimulated expression of another signal, parathyroid hormone-related protein (PTHrP), at the tip of the growth plate¹². Because the cells that express PTHrP are found some distance away from those expressing Ihh, mutations that compromise the diffusion of Ihh might be expected to cause reduction in bone length — exactly as

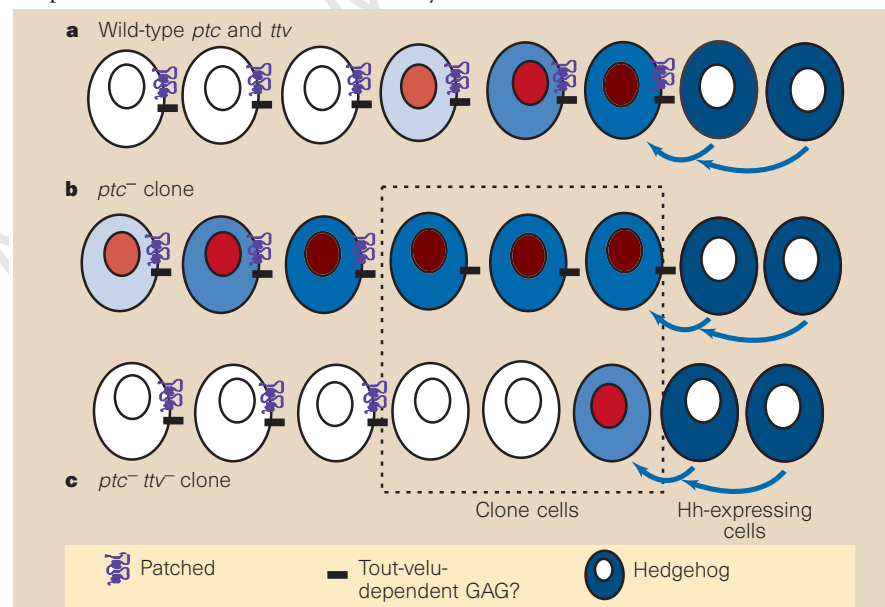


Figure 1 Diffusion of the Hedgehog (Hh) protein in *Drosophila* is regulated by the Patched (Ptc) and Tout-velu (Ttv) proteins. *Ttv* has just been cloned by Bellaïche *et al.*² and is homologous to *EXT-1*, which encodes a protein that is implicated in glycosaminoglycan (GAG) synthesis¹⁰. a, In wild-type fruitflies, Ptc and Ttv have opposing effects on the movement of Hh. Diffusion is constrained because Hh is sequestered by its transmembrane receptor, Ptc. The activity of Ttv, by contrast, helps to mediate Hh movement, presumably through its effects on synthesis of an unidentified cell-surface GAG. b, In the absence of Ptc, Hh can diffuse freely. c, In the absence of both Ptc and Ttv, Hh cannot diffuse far. Presumably, although there is no Ptc to sequester Hh, the Ttv-dependent GAG is not present to facilitate its diffusion. Filled nuclei indicate *hh* target-gene activation.

is seen in patients with multiple exostoses. Paradoxically, however, there is no evidence to date that *Ihh* signals directly across the growth plate to the PTHrP-expressing cells; rather, available data indicate that *Ihh*-expressing cells signal to the adjacent perichondrium, which in turn signals to the PTHrP cells, probably through a bone morphogenetic protein¹³.

Because affected people develop bone tumours, *EXT-1* and *EXT-2* could be tumour-suppressor genes, mutation of both alleles of either locus being necessary for tumour formation³. This interpretation fits with the aetiology of the disease, but the insights by Bellaïche and colleagues² into the probable role of the *EXT* proteins, based on their analysis of *Drosophila* *ttv*, raise a question mark over such a model. If either *EXT* protein is required for diffusion of *Ihh*, its elimination might be expected to cause premature exit of chondrocytes from the cell cycle, rather than their unrestrained proliferation, which presumably results in exostoses. In fact, analysis¹⁴ of *EXT-1*-associated malignancies argues against the simple tumour-suppressor model — loss of het-

erozygosity in an exostoses-derived chondrosarcoma is due to elimination of the mutant, not the wild-type, *EXT-1* allele. So, complete loss of function of *EXT-1* is not a prerequisite for tumour formation. There is much still to be learned about this interesting condition, but as the work of Bellaïche *et al.* has shown, further clues from *Drosophila* will doubtless help. □

Philip W. Ingham is in the Developmental Genetics Programme, The Krebs Institute, University of Sheffield, Firth Court, Sheffield S10 2TN, UK.

e-mail: p.w.ingham@sheffield.ac.uk

1. Ingham, P. W. *EMBO J.* **17**, 3505–3511 (1998).
2. Bellaïche, Y., The, I. & Perrimon, N. *Nature* **394**, 85–88 (1998).
3. Ahn, J. *et al.* *Nature Genet.* **11**, 137–143 (1995).
4. Stickens, D. *et al.* *Nature Genet.* **14**, 25–32 (1996).
5. Porter, J., Young, K. & Beachy, P. *Science* **274**, 255–259 (1996).
6. Chen, Y. & Struhl, G. *Cell* **87**, 553–563 (1996).
7. Perrimon, N., Lanjuin, A., Arnold, C. & Noll, E. *Genetics* **144**, 1681–1692 (1997).
8. Taylor, A. M., Nakano, Y., Mohler, J. & Ingham, P. W. *Mech. Dev.* **43**, 89–96 (1993).
9. Tabata, T. & Kornberg, T. B. *Cell* **76**, 89–102 (1994).
10. McCormick, C. *et al.* *Nature Genet.* **19**, 158–161 (1998).
11. Solomon, L. J. *Bone Joint Surg.* **45B**, 292–304 (1963).
12. Vortkamp, A. *et al.* *Science* **273**, 613–622 (1996).
13. Zou, H., Wieser, R., Massag, e, J. & Niswander, L. *Genes Dev.* **11**, 2191–2203 (1997).
14. Hecht, J. *et al.* *Am. J. Hum. Genet.* **60**, 80–86 (1997).

Extragalactic astronomy

Intergalactic pollution

J. Michael Shull

Astronomers are looking for the first generation of stars and galaxies, and for their imprint on the chemistry of the Universe. It is a quest with a frontier spirit about it. Cosmological nucleosynthesis models¹ predict that the elements to emerge from the Big-Bang fireball were hydrogen, helium and traces of deuterium and lithium. Heavier elements (called ‘metals’ by astronomers) such as carbon, oxygen, silicon and iron, only formed much later, inside stars. But on page 44 of this issue,

Cowie and Songaila² discuss new data from the High Resolution Spectrometer on the Keck 10-metre telescope that hint that the low-density intergalactic medium (IGM) is far more metal-rich than expected. Where could these heavy elements have come from?

Massive stars inject metal-enriched gas into interstellar space when they undergo supernova explosions. Pressure-driven outflows from galaxies then blow this debris into the IGM (Fig. 1). That should affect the chemistry and temperature around early quasars and star-forming objects, but regions of the IGM far from these objects should have little heavy-element ‘pollution’. The best place to look for pristine gas is in voids between galaxy clusters: hydrodynamical simulations^{3,4} predict that gravitational instabilities created large filamentary structures in the IGM. Massive galaxies formed in the dense filaments of gas and dark matter, and low-density gas filled the voids between the filaments. Large-scale redshift surveys bear out this picture (Fig. 2).

Because most star formation and metal production probably occurred in the densest regions of the Universe, the low-density voids between galaxies should be the cosmological frontier, with the lowest metal abundance^{5,6}. Indeed, until a few years ago, all weakly absorbing hydrogen clouds were thought to be pristine matter⁷ consisting only of primordial hydrogen and helium. The strongest hydrogen absorbers were later shown^{8,9} to contain traces of carbon and silicon ions, and were taken as indirect evidence for the escape of enriched gas from ancient galaxies.

To test this further, Cowie and Songaila took the spectra of six quasars at redshifts $z > 3$, looking for weak absorption lines caused by the gas along our line of sight — the lines from triply ionized carbon (C IV) and neutral hydrogen (H I). From a pixel-by-pixel comparison of the carbon and hydrogen absorption, they find that the C IV/H I abundance ratio is roughly the same in all gas clouds, down to the weakest levels that can be

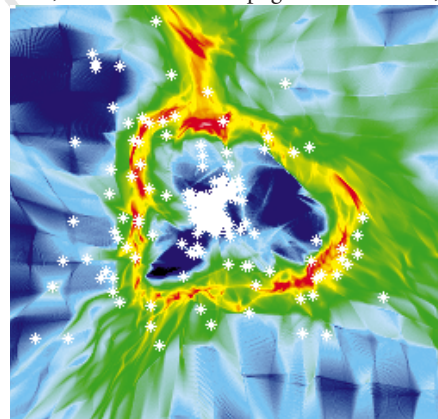


Figure 1 Simulation⁶ of the gas density around a star-forming galaxy at redshift $z = 9$. The box size is 200 kpc (for a Hubble constant of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$); white symbols are stars. Heavy elements would be dispersed by the outflows predicted here.

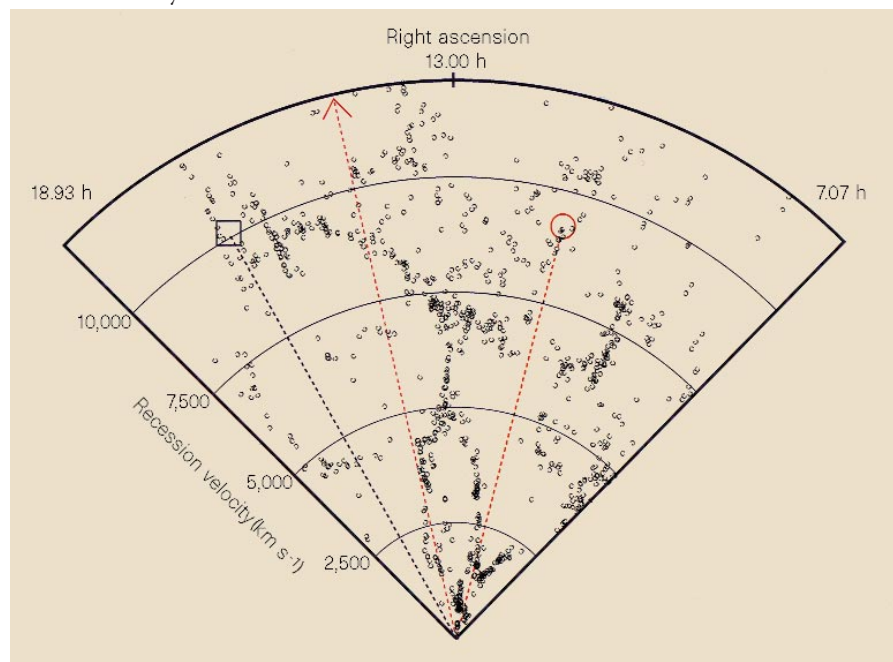


Figure 2 A slice of the local galaxy distribution. Distinct ‘walls’ and ‘voids’ can be seen, on a scale of up to about 50 megaparsecs (equivalent to more than $3,000 \text{ km s}^{-1}$). Sightlines to nearby quasars, used as probes of the intervening gas¹³, are also shown.

measured (corresponding to the most tenuous regions). Earlier analyses of C IV/H I ratios^{10,11} also yielded a value, ~ 0.003 times that in the Sun, that is approximately constant with H I column density for the higher-density clouds.

This seems to go against the theory that void gas should be pristine. What could explain its enrichment? At $z = 3$ the Universe is about two billion years old, and in that time heavy elements expelled from early star-forming galaxies at 300 km s^{-1} would only have moved a small fraction of the 5–10-megaparsec size of the galaxy voids (Fig. 2). The first star-forming objects might have generated 'super-winds', travelling at more than $1,000 \text{ km s}^{-1}$, that took metals to greater distances and thereby smoothed out the metallicity contrast between galaxy walls and voids. Alternatively, the voids may be inhabited by dwarf galaxies that easily shed their gas and metals following a flurry of massive-star formation and supernovae, and then fade. Or could there have been a widely distributed generation of 'population III' early stars that formed before the segregation into voids and filaments, and polluted the Universe to a nearly uniform metallicity?

But perhaps this is all a false alarm. Using a different statistical technique, and adding together about 300 weak C IV lines from gas on our lines of sight towards nine quasars, Lu *et al.*¹² see a lower carbon abundance in the low-matter-density IGM than in the high-density areas. They believe that their data can rule out the population-III idea.

It is tempting to presume that one of the groups is wrong. But it could be that the IGM has a wide distribution of heavy-element abundances that neither group has adequately characterized. Like blind men describing an elephant, different observers may be prob-

ing different parts of the IGM density and metallicity distribution, and failing to get a good picture of the whole. Each group should look carefully at the other's analysis techniques (they have two quasars in common) and continue the debate. Depending on how it is resolved, astronomers may need to revise their views on the first epoch of star formation and the physical mechanisms that return gas to intergalactic space.

The new Keck results push the limits of spectroscopy, with integration times of 4 to 11 hours, reaching 100:1 signal-to-noise ratios towards 16–17th-magnitude quasars. They highlight the need to characterize detector noise and to devise statistical techniques that dig out weak features. New 10-metre telescopes will soon be joining the quest, and astronomers are planning new infrared, X-ray and ultraviolet instruments to search for other traces of the first generations of star-forming galaxies. □

J. Michael Shull is in the Department of Astrophysical and Planetary Sciences, University of Colorado at Boulder, Boulder, Colorado 80309, USA. e-mail: mshull@casa.colorado.edu

1. Copi, C. J., Schramm, D. N. & Turner, M. S. *Science* **267**, 192–199 (1995).
2. Cowie, L. L. & Songaila, A. *Nature* **394**, 44–46 (1998).
3. Zhang, Y., Meiksin, A., Anninos, P. & Norman, M. L. *Astrophys. J.* **485**, 496–516 (1997).
4. Hernquist, L., Katz, N., Weinberg, D. H. & Miralda-Escudé, J. *Astrophys. J.* **471**, 582–616 (1996).
5. Gnedin, N. Yu. & Ostriker, J. P. *Astrophys. J.* **486**, 581–598 (1996).
6. Gnedin, N. Yu. *Mon. Not. R. Astron. Soc.* **294**, 407–421 (1998).
7. Sargent, W. L. W., Young, P. J., Boksenberg, A. & Tytler, D. *Astrophys. J. Suppl. Ser.* **42**, 41–81 (1980).
8. Cowie, L. L., Songaila, A., Kim, T.-S. & Hu, E. M. *Astron. J.* **109**, 1522–1530 (1995).
9. Tytler, D. *et al.* in *QSO Absorption Lines* (ed. Meylan, G.) 289–298 (Springer, Berlin, 1995).
10. Songaila, A. & Cowie, L. L. *Astron. J.* **112**, 335–351 (1996).
11. Giroux, M. L. & Shull, J. M. *Astron. J.* **113**, 1505–1513 (1997).
12. Lu, L., Sargent, W. L. W., Barlow, T. A. & Rauch, M. *Astron. J.* (submitted); preprint astro-ph/9802189 on xxx.lanl.gov (1998).
13. Shull, J. M., Stocke, J. T. & Penton, S. *Astron. J.* **111**, 72–77 (1996).

Synaptic transmission

The two faces of glutamate

Jean-Philippe Pin

Most fast, excitatory synapses in the brain use the common neurotransmitter glutamate, which activates both ionotropic and metabotropic glutamate receptors (iGluRs and mGluRs). Because mGluRs are located on both pre- and postsynaptic elements, they are important in regulating the strength and frequency of synaptic firing, including the long-term modification of synaptic strength. Surprisingly few effects resulting from activation of the mGluRs by synaptically released glutamate have been identified^{1,2} but, on page 78 of this issue, Fiorillo and Williams³ reveal a new one. They have found that one mGluR that was assumed to be mainly excitatory can generate a large inhibitory action upon release of glutamate.

Inhibition by glutamate (hyperpolarization) has been known for some time in invertebrates. At the neuromuscular junction of insects, glutamate excites nerve cells to induce contraction — just as acetylcholine does for contraction of our own muscles. However, extrasynaptic glutamate receptors located on insect muscles have a hyperpolarizing action that results from activation of a chloride channel⁴. In mammals, glutamate is also the transmitter in a well-known inhibitory synapse in the retina⁵. This synapse — between the photoreceptor cell and the ON bipolar cell — is responsible for the signal recorded from the retina when light is switched on. In the dark, constant release of glutamate by the photoreceptor cells inhibits the activity of the ON bipolar cells by activat-

ing an mGluR. This receptor, in turn, inhibits the activity of a cation channel.

The inhibitory action of glutamate can also be mediated by a chloride channel that is associated with a high-affinity glutamate transporter⁶, or by activation of potassium channels^{7,8}. Accordingly, synaptically released glutamate in the brain was expected to show inhibitory actions, and Fiorillo and Williams³ now report the first example of this. They have characterized a mGluR (mGluR1) which, by activating phospholipase C (PLC; the enzyme that catalyses production of the second messengers inositol-1,4,5-trisphosphate and diacylglycerol), activates an apamine-sensitive potassium channel (Fig. 1, overleaf).

Why has this inhibitory action of glutamate rarely been detected before? It could be that it is a specific property of only a few glutamatergic synapses, although that in turn may seem to occur simply because the inhibitory effect can be observed only under particular conditions. For example, Fiorillo and Williams did not observe the effect when selective agonists for mGluR1 were applied in the bath (slow application) — instead, they found a very slowly developing excitation (depolarization). This agrees with observations by others^{2,9} and with the general idea that PLC-coupled receptors exert potentiating effects by inhibiting potassium channels or activating non-selective cation channels (Fig. 1).

Fiorillo and Williams also found mGluR1-mediated excitation upon synaptic release of glutamate, but after the inhibitory phase. Only fast application of mGluR agonists can generate the inhibitory action (Fig. 1), and the authors propose that the intracellular cascade responsible for the inhibitory desensitizes rapidly. In other words, all of the receptors have to be activated at the same time to generate this response. So, these findings indicate that the same receptor can mediate either excitatory or inhibitory effects depending on the kinetics of its stimulation.

A similar switch between positive and negative effects mediated by a PLC-coupled mGluR has also been reported at the presynaptic level. Sanchez-Prieto and colleagues used biochemical methods to show¹⁰ that activation of a presynaptic PLC-coupled mGluR could potentiate the release of glutamate. More recently, these authors have demonstrated¹¹ that the potentiating effect that they previously described is turned into an inhibitory effect if the PLC response mediated by this receptor is desensitized. They did this by first activating the receptor (or by constantly activating it) with a low concentration of glutamate. So, as at the postsynaptic side, depending on how the presynaptic PLC-coupled mGluRs are activated, they will exert either an inhibitory or a potentiating effect. This may explain why a

NATURE

100 YEARS AGO

"Liquid Hydrogen" – I observe with some amusement that you still allow Mr Hampson to embellish your columns with vain repetitions of accusations which he was compelled to withdraw when he met me face to face at the meeting of the Society of Chemical Industry. ... Mr Hampson must be a singularly dull person if he fails to appreciate the magnitude of the draft he makes upon the credulity of the world. He asks men of the world to believe that he, being convinced of the general dishonesty of Royal Institution methods, and being in possession of a novel and valuable invention, fully completed but not protected by patent, came unbidden and unsought to reveal all the details to a man whom he knew to be my assistant. ... If all that Mr. Hampson wants is "recognition in historical or explanatory works" of his claim to be the inventor of a general claim to intensive refrigeration, he will find Solvay, Dr. Linde, and Prof. Onnes obstacles quite as serious as myself. Further, this attempt to justify going behind my back in his relations with a member of the staff of the Royal Institution, is a too transparent subterfuge to require further comment. – James Dewar

From *Nature* 30 June 1898.

50 YEARS AGO

... Many of us in Britain can realize only too well the strains to which a German scientist of integrity has been subjected; the conflict between the desire to stay and save from the Nazis something of German culture and the desire to escape to a free country; the reluctance to be an "armourer of Himmler and of Auschwitz", and the desire to strike back at the destroyers of Hamburg and of Dresden. In a few years time, when passions have cooled a little, the welcome that a German scientist will find among his colleagues in the West should depend very little on the extent to which, under the stress of war, he worked on weapons. It is to be hoped that it will depend far more on his earlier record, on the opposition that he offered within and outside his university, at the time when opposition was still possible, to the steps that stifled freedom in Germany and made war inevitable. – N. F. Mott

From *Nature* 3 July 1948.

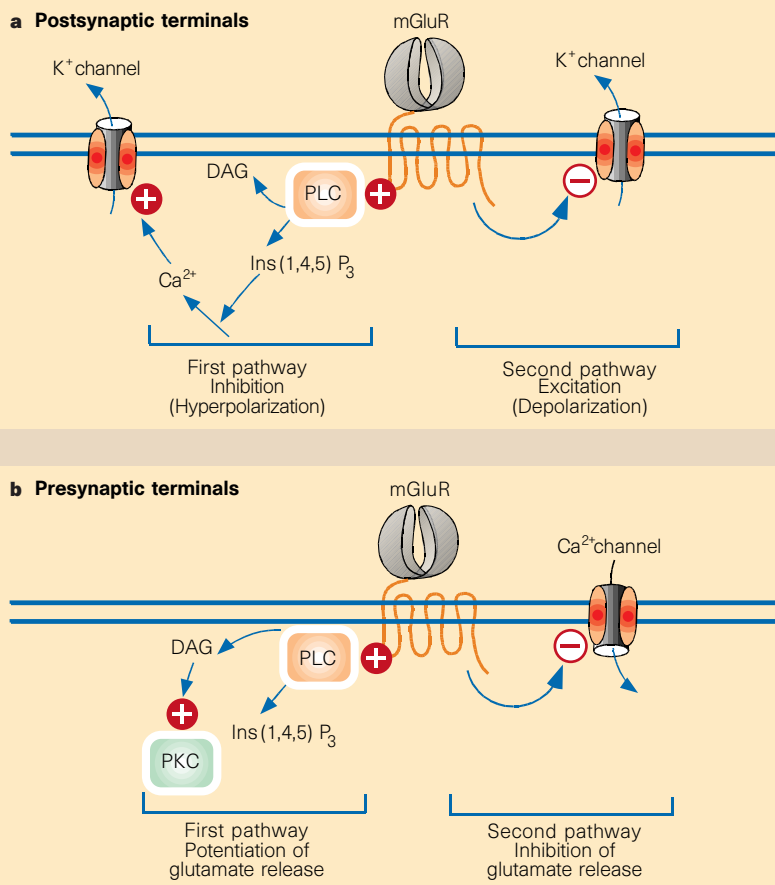


Figure 1 Effects of the neurotransmitter glutamate on synaptic transmission. Glutamate mediates fast, excitatory synaptic transmission by activating specific receptor-channels. It also activates metabotropic receptors (mGluR), which modulate this synaptic transmission. **a**, Fiorillo and Williams³ have found that fast activation of an mGluR sub-type by synaptically released glutamate mediates a slow inhibition of the postsynaptic cell (first pathway). In contrast, a slow activation of the same receptor leads to the opposite effect — a slowly developing excitation (second pathway). **b**, If the phospholipase C (PLC) pathway is not desensitized, the same type of mGluR located on presynaptic terminals can potentiate the release of glutamate (first pathway)¹¹. Similarly, if the PLC pathway is desensitized, glutamate release from presynaptic terminals will be inhibited (second pathway). These findings^{3,11} indicate that the same receptor type can mediate either positive or negative effects depending on how it is activated.

presynaptic potentiating effect of mGluRs has rarely been observed when the activity of excitatory synapses is recorded — instead, an inhibitory effect of presynaptic PLC-coupled mGluRs has been found¹².

We don't know what determines this switch in activity of the mGluR but, interestingly, the β_2 -adrenergic receptor switches from stimulating to inhibiting adenylyl cyclase after it has been desensitized by protein kinase A (ref. 13). So, opposite effects from a single type of receptor might be expected for other guanine-nucleotide-binding (G)-protein-coupled receptors.

Glutamate also activates NMDA (N-methyl-D-aspartate) receptors, and many glutamatergic synapses have peculiar properties because of this. Activated NMDA receptors allow the entry of Ca^{2+} , which induces biochemical changes in the post-

synaptic element leading to modifications of the synaptic strength. However, glutamate-mediated activation of NMDA receptors is effective only if associated with a depolarization, so it will be more difficult to activate them during the inhibitory phase that is generated by mGluRs. Conversely, it will be easier to activate them later on, during the excitatory phase. It is therefore not surprising that the delay that separates the arrival of two distinct excitatory inputs is crucial for the induction of synaptic plasticity. Moreover, the results of Fiorillo and Williams³ may help us to understand the apparent contradictory and complex effects of PLC-coupled mGluRs in synaptic plasticity¹⁴, revealing a new level of complexity in the functioning of glutamatergic synapses. □

Jean-Philippe Pin is at the CNRS-UPR 9023, Mécanismes Moléculaires des Communications

Cellulaires, CCIPE, 141 rue de la Cardonille, 34094, Montpellier Cedex 5, France.

e-mail: pin@ccipe.montp.inserm.fr

1. Scanziani, M., Salin, P. A., Vogt, K. E., Malenka, R. C. & Nicoll, R. A. *Nature* **385**, 630–634 (1997).
2. Batchelor, A. M. & Garthwaite, J. *Nature* **385**, 74–77 (1997).
3. Fiorillo, C. D. & Williams, J. T. *Nature* **394**, 78–82 (1998).
4. Cull-Candy, S. G. & Usherwood, P. N. R. *Nature* **246**, 62–64 (1973).
5. Masu, M. *et al.* *Cell* **80**, 757–765 (1995).
6. Fairman, W. A., Vanderberg, R. J., Arriza, J. L., Kavanaugh, M. P. & Amara, S. G. *Nature* **375**, 599–603 (1995).
7. Rannin, D. G., Holmes, K. H. & Shinnick-Gallagher, P.

- J. Neurosci.* **14**, 7208–7220 (1994).
8. Fagni, L., Bossu, J.-L. & Bockaert, J. *Eur. J. Neurosci.* **3**, 778–789 (1991).
9. Charpak, S., Gähwiler, B. H., Do, K. Q. & Knöpfel, T. *Nature* **347**, 765–767 (1990).
10. Herrero, I., Miras-Portugal, M. T. & Sanchez-Prieto, J. *Nature* **360**, 163–166 (1992).
11. Herrero, I., Miras-Portugal, M. T. & Sanchez-Prieto, J. *J. Biol. Chem.* **273**, 1951–1958 (1998).
12. Gereau, R. W. & Conn, P. J. *J. Neurosci.* **15**, 6879–6889 (1995).
13. Daaka, Y., Luttrell, L. M. & Lefkowitz, R. J. *Nature* **390**, 88–91 (1997).
14. Bortolotto, Z. A., Bashir, Z. I., Davies, C. H. & Collingridge, G. L. *Nature* **368**, 740–743 (1994).

Geophysics

Ice on the fast track

Charles R. Bentley

The West Antarctic Ice Sheet (WAIS) has a purported ability to shrink rapidly, or 'collapse', causing a disastrous rise in sea level of some five metres¹. But is it likely to do so? To answer that, we must first understand the fast-flowing ice streams that stud its perimeter. Ice streams are glaciers bounded at the sides by nearly stationary ice, rather than by rock walls, and they are the main route to the sea for ice in the WAIS. What controls their locations? Bell *et al.*² and Anandakrishnan *et al.*³ (on pages 58 and 62 of this issue) propose a partial answer: ice streams in the WAIS exist only over sedimentary basins in the underlying ground. A similar association of fast flow with underlying sediment has been proposed for the Laurentide Ice Sheet in Canada during the last ice age⁴.

Mean sea level around the world is rising at a modest 0.2 metres per century — easily measurable, and even observable by people who have lived along the coast for a long time and have accurate memories, but

hardly catastrophic. At the moment, the masses of the Antarctic and Greenland ice sheets, vast repositories of water held out of the ocean, remain nearly constant. In contrast, while the Laurentide and other Northern Hemisphere ice sheets were melting away at the end of the last ice age, sea level rose by several metres per century at times⁵. Such rates of rise now would be disastrous for coastal populations. So can the present ice sheets shrink that fast? If so, can we predict when and how rapidly the shrinkage actually will take place? Opinions differ widely^{6,7}.

At present the WAIS is surrounded by mountains and vast, floating ice shelves. It is the most likely ice sheet to shrink rapidly, because it rests on a bed that is mostly far below sea level⁸. Such a low-lying bed could make the 'grounding line' unstable (the grounding line is the boundary between the interior, grounded ice and the floating ice shelves; Fig. 1) because the thick ice shelf at

the grounding line should spread out and thin more rapidly than the ice can be replaced from the interior. Thus the ice thins and the grounding line retreats, perhaps transforming the massive interior ice sheet into a gigantic ice shelf. That would be enough to raise sea level — melting of the ice would not be necessary, because floating ice already displaces its weight in water. But to know whether or when such an instability will occur, the effects of ice streams must be taken into account — for one thing, they may supply ice to the grounding line fast enough to prevent grounding-line retreat.

In the United States Antarctic Program, research has concentrated on the Ross ice streams, the half-dozen streams up to 150 km across that flow into the Ross Ice Shelf. Despite surface slopes that are gentle compared with those on the surrounding, slowly moving ice, the ice streams flow a hundred times faster — hundreds of metres per year instead of metres per year. This is contrary to expectation, because it is the component of gravity along the ice surface that drives any glacier forward. But the ice streams are known to rest on extremely slippery beds — water-saturated (diluted), highly porous, fine-grained sediments overlain by a film of water — that only weakly restrain the ice sliding over them. The restraint is too weak to balance the gravitational driving force on the ice stream: the balance is maintained by a combination of shear stresses along their sides, where the streams abut the neighbouring slow-moving ice, and friction at 'sticky spots' on the bed, where the efficient lubrication is apparently absent¹.

This complex picture still does not predict where the ice streams develop within an ice sheet. According to Bell *et al.*² and Anandakrishnan *et al.*³, the answer is, at least in one place, that control lies in the subglacial geology. They explored the geology with a well-chosen combination of measurements: Bell *et al.* used satellite data on gravity, magnetism and subglacial topography (using ice-penetrating radar); Anandakrishnan *et al.* used seismic data from controlled explosions on the ice surface. They have discovered that the upstream end of one ice stream is co-extensive with a sedimentary basin.

The implication they draw is that the underlying sedimentary rocks are the source, through erosion, of the lubricating layer of soft diluted sediment immediately below the ice, and therefore that an ice stream can only exist where there are sedimentary rocks thick enough (more than 100 m; ref. 2) to yield a long-term supply of diluted sediment.

This interpretation is intriguing, but speculative. For instance, the sedimentary basin under the upstream end of the ice stream is a deep valley. Could this mean that the association between the ice stream and the sediment is just coincidence — are both

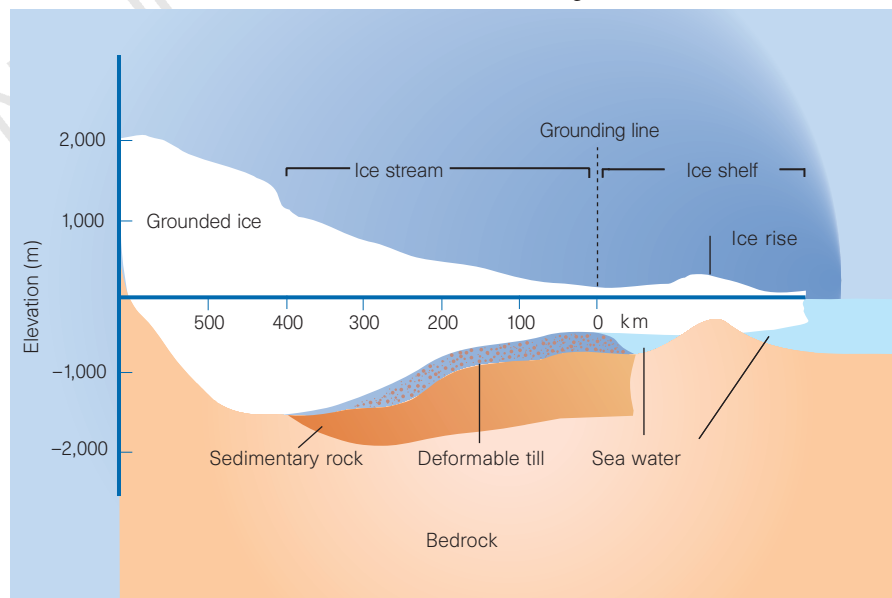


Figure 1 Cross-section through an ice sheet. Because the bed lies far below sea level, this ice sheet may be unstable — but we don't yet know whether that is so. The roles of the slippery sediment (glacial till) underlying the ice and the deeper sedimentary rock, which may supply the till, could be crucial. (Adapted from ref. 1.)

of them there because the valley is there? But then again, some portions of the Ross ice streams overlies beds that are not in deep valleys; in other places there are subglacial troughs but the ice streams lie only partially over them⁹ — so is the distribution of sedimentary rocks perhaps the control in those places? Still elsewhere there are only thin sedimentary rocks under an ice stream, but a sedimentary basin under the adjacent, nearly stagnant ice¹⁰. What is the control there? The erosion rate under dilated sediment is probably less than 0.1 mm yr⁻¹ (S. Tulaczyk, personal communication); why, then, must the sedimentary rock column be thick compared with 100 m, which would already be enough to supply sediment for over a million years? Do the different Ross ice streams perhaps differ in their basic dynamic control? Most important of all, what does all this do to ice-sheet stability?

Whatever the answers, these papers show that subglacial geology should be added to the complex mix of factors considered in

explaining the behaviour of the WAIS and predicting its future. Just how important that geology is can perhaps be answered with more geophysical studies. But for the time being, glaciologists, little as they may relish having to consider yet another complication to the ice-stream problem, appear to have no choice but to do so. □

Charles R. Bentley is at the Geophysical and Polar Research Center, University of Wisconsin, 1215 West Dayton Street, Madison, Wisconsin 53706, USA.

e-mail: bentley@geology.wisc.edu

1. Oppenheimer, M. *Nature* **393**, 325–332 (1998).
2. Bell, R. E. *et al.* *Nature* **394**, 58–62 (1998).
3. Anandakrishnan, S., Blankenship, D. D., Alley, R. B. & Stoffa, P. L. *Nature* **394**, 62–65 (1998).
4. Clark, P. U., Licciardi, J. M., MacAyeal, D. R. & Jenson, J. W. *Geology* **24**, 679–682 (1996).
5. Bard, E. *et al.* *Nature* **382**, 241–246 (1996).
6. Bentley, C. R. *Science* **275**, 1077–1078 (1997).
7. Bindshadler, R. *Ann. Glaciol.* **24**, 409–414 (1997).
8. Alley, R. B. & Whillans, I. M. *Science* **254**, 959–963 (1991).
9. Shabtaie, S., Whillans, I. M. & Bentley, C. R. *J. Geophys. Res.* **92**, 8865–8883 (1987).
10. Munson, C. G. & Bentley, C. R. in *Recent Progress in Antarctic Earth Science* (eds Yoshida, Y. *et al.*) 507–514 (Terrapub, Tokyo, 1992).

quasiparticles) have acquired very large effective masses, often hundreds of times that of a free electron. These large masses arise from the interaction of the conduction electrons with localized (generally f-shell) magnetic moments in the materials. When this interaction favours antiparallel alignment of the local and conduction-electron spins, it is possible for the conduction electrons to screen the local magnetic moment, resulting in a non-magnetic state at temperatures below that characterizing the strength of this interaction. The conduction electrons acquire mass from the local moments they screen.

The underlying magnetic provenance of this heavy-fermion state might appear to make them unlikely candidates for superconductors, as magnetic interactions tend to break antiparallel-spin Cooper pairs. But some of them become superconducting in spite of this, albeit at temperatures generally of the order 1 K, and so there has been speculation from the first discovery of superconductivity in such a material (CeCu₂Si₂) as to whether antiparallel spin pairing can actually be mediated by a magnetic interaction.

In the experiments of Mathur *et al.*, a magnetically ordered state of cerium 4f-local moments competes with a heavy-fermion state that has all these moments compensated by conduction electrons. Pressure increases the conduction-electron/4f-moment coupling in cerium materials, favouring the heavy-fermion state. But in between low-temperature magnetic order and heavy-fermion behaviour, superconductivity emerges. Mathur *et al.* argue that superconductivity arises here from magnetically mediated pairing — what they call

Superconductivity

The ghost of magnetism

Zachary Fisk and David Pines

In 'conventional' superconductivity, electrons form superconducting Cooper pairs through an attraction mediated by lattice vibrations. Are any other pairing mechanisms possible? As well as having intrinsic physical interest, such mechanisms might allow much higher transition temperatures. But the phenomena of superconductivity are remarkably independent of the actual mechanism, so

it is only through indirect methods that we can uncover the forces at work. Nevertheless, Mathur *et al.* (reporting on page 39 of this issue¹) have inferred that a magnetic pairing mechanism operates in two materials.

The subjects of this study are two heavy-fermion superconductors. These are metallic materials in which the conduction electrons (or 'dressed' collective states called

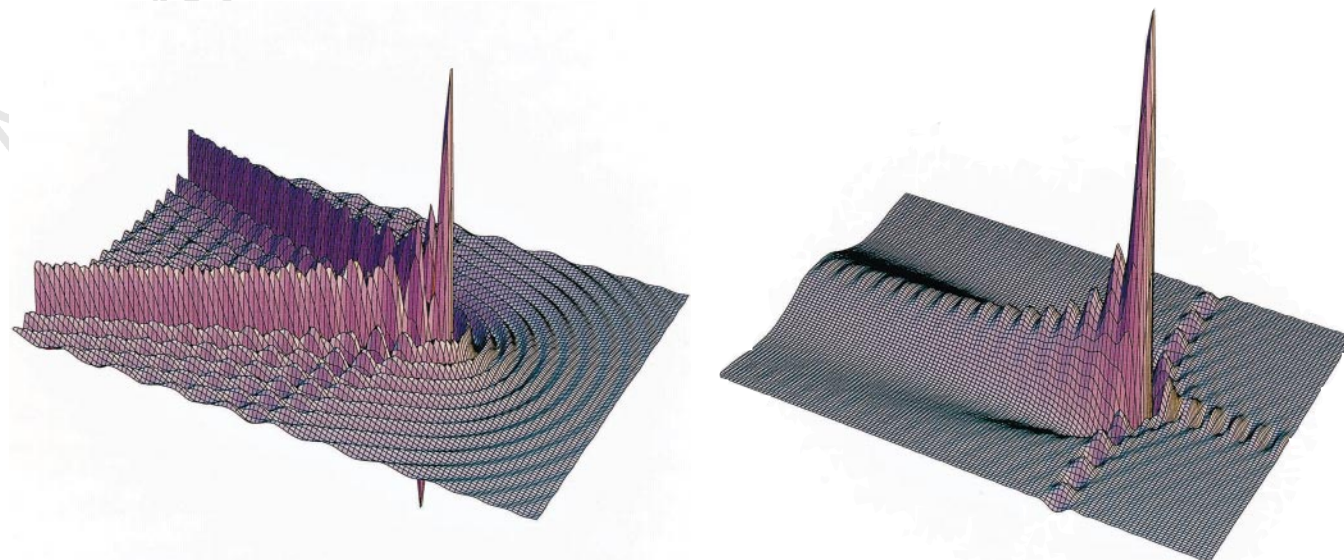


Figure 1 Sources of superconductivity. Left, the interaction potential (in one plane) generated by a test charge moving rapidly through a metal. A repulsive core is followed by a wake, and, crucially, an attractive well caused by lattice vibrations, which can lead to charge-carrier pairing and conventional superconductivity. Right, the spin of the charge carrier also generates a magnetization shock wave. (In a

nearly antiferromagnetic metal, this is the envelope of a rapidly varying modulation.) The interaction potential should have a similar form, but with a sign that depends on the relative orientation of the interacting spins. Near to a state of long-range magnetic order, this spin-spin interaction may dominate, and so lead to novel forms of superconductivity.

'magnetic glue'. One can also think of magnetism and superconductivity as different manifestations of the same basic interaction: spin-density waves, generated by moving charge carriers (Fig. 1), fail to propagate, and instead serve to bind these particles into pairs. The characteristic energy of this binding corresponds to the transition temperatures observed in the real materials CePdSi_2 and CeIn_3 . The participation of quasiparticles in a nascent spin-density wave (the ghost of magnetism) is inevitably accompanied by a reduction in the quasiparticle spectral strength at low energy.

Mathur *et al.* find a second signature of a magnetic interaction between quasiparticles: the electrical resistivity varies as a power of temperature different from the T^2 -dependence of a conventional Fermi liquid (an electron gas with weak interactions). Similar interactions have been postulated in the high-temperature cuprate superconductors as a possible explanation for the so-called pseudogap — a region of low density in quasiparticle states². It would be interesting to know if there is a comparable effect in the materials studied by Mathur *et al.* Something like this is seen³ in the optical conductivity of the heavy-fermion superconductor UPt_3 .

It would not be surprising if more candidates for a purely electronic mechanism for superconductivity turn up in other members

of the strongly correlated electron family, as the strong interactions that characterize these systems make them prime candidates for the presence of 'failed' spin- or charge-density waves, either of which might bring about superconductivity.

Finding that a superconducting phase reflects the ghost of magnetism prompts the converse question: in what magnetic systems can we see the ghost of superconductivity? It seems to be there in the magnetic properties of the insulating layered cuprate material $\text{Sr}_2\text{CuO}_2\text{Cl}_2$, taken close to the superconducting transition⁴; and it is certainly inherent to the $\text{SO}(5)$ group theoretical approach that attempts to unify magnetism and superconductivity in the cuprates⁵. The study of this new physical phenomenon, whose offspring are magnetic and superconducting states, now has some direction. □

Zachary Fisk is in the Department of Physics, Florida State University, Tallahassee, Florida 32306, USA.

e-mail: fisk@magnet.fsu.edu

David Pines is in the Department of Physics, University of Illinois, 1110 West Green Street, Urbana, Illinois 61801-3080, USA.

e-mail: d-pines@uiuc.edu

1. Mathur, N. D. *et al.* *Nature* **394**, 39–43 (1998).

2. Schmalian, J. *et al.* *Phys. Rev. Lett.* **80**, 3839–3842 (1998).

3. Donovan, S. *et al.* *Phys. Rev. Lett.* **79**, 1401–1404 (1997).

4. Kim, C. *et al.* *Phys. Rev. Lett.* **80**, 4245–4248 (1998).

5. Zhang, S.-C. *Science* **275**, 1089–1096 (1997).

Signal transduction

What goes up must come down

David R. H. Evans and Brian A. Hemmings

Eukaryotic cells can adjust their metabolism, growth and differentiation in response to hormones, growth factors, neurotransmitters and stress by reversible protein phosphorylation — the addition and removal of phosphate groups. To ensure the efficacy of intracellular signals conveyed in this manner, opposing protein

kinase and protein phosphatase enzymes (which add and remove phosphate groups, respectively) have evolved a high degree of substrate specificity, and systems for precise intracellular targeting^{1,2}. Reports by Westphal *et al.*³ and Camps *et al.*⁴ in *Science* now highlight an additional molecular device for the feedback regulation of kinase signalling

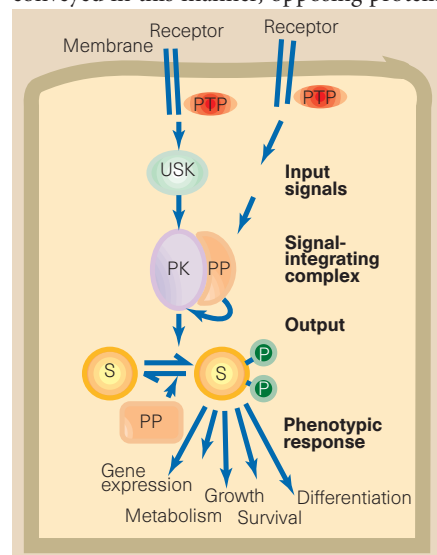


Figure 1 Signalling cascade responding to 'on' and 'off' signals. Signals received at the cell membrane stimulate the activation of a signal-integrating complex, which incorporates both a protein kinase and a protein phosphatase. Specific examples of such complexes are described by Westphal *et al.*³ and Camps *et al.*⁴. Signals integrated by the complex are transmitted to the substrate, eliciting a phenotypic response. PTP, protein tyrosine phosphatase; USK, upstream kinase; PK, protein kinase; PP, protein serine/threonine phosphatase or dual-specificity phosphatase (for serine/threonine and tyrosine); S, substrate; P, phosphate group.

pathways. Here, a protein kinase, which is itself regulated by protein phosphorylation, becomes the substrate of a specific protein phosphatase in a self-moderating signalling complex.

Westphal *et al.*³ describe an interaction between the Ca^{2+} -calmodulin-dependent protein kinase IV (CaMKIV) and the ubiquitous and evolutionarily conserved protein phosphatase 2A (PP2A). CaMKIV is activated in a Ca^{2+} -dependent manner involving phosphorylation by the upstream kinase Ca^{2+} -calmodulin-dependent protein kinase kinase (CaMKK). Activation is transient and reversible, even in the presence of high levels of intracellular Ca^{2+} . The authors found that PP2A binds to the region of CaMKIV that encompasses the kinase catalytic domain, but that this binding is independent of CaMKK-mediated phosphorylation. Moreover, although CaMKK-driven activation of CaMKIV is reversed by PP2A, the enzymatic activity of neither CaMKIV nor PP2A is required for their physical interaction, revealing a stable interaction.

The tight regulation of CaMKIV phosphorylation by PP2A in a constitutive signalling complex explains why CaMKIV is only transiently activated by CaMKK in response to the concentration of intracellular Ca^{2+} . The result also provides a model for understanding the dynamics of other kinase signalling pathways — clearly, in a kinase/phosphatase complex, signals transmitted by the kinase may be adjusted promptly to achieve an appropriate cellular response. Nevertheless, questions remain regarding the self-regulation of CaMKIV/PP2A and similar complexes. PP2A binds to CaMKIV as a trimeric holoenzyme consisting of a catalytic subunit C, and two regulatory subunits, A and B. What are the contacts made by CaMKIV with the various PP2A subunits? And does, for example, the AC core dimer of PP2A serve as an obligatory platform for the binding of CaMKIV, as it does for the binding of its own B subunit? Furthermore, does PP2A dephosphorylate CaMKIV constitutively, being outpaced by CaMKK only when the concentration of intracellular Ca^{2+} is high? Although attractive, this model is overly simple because PP2AC may be regulated by many post-translational modifications, including phosphorylation and methylation, and it is targeted by specific inhibitor proteins⁵. Moreover, signalling through the TOR (target of rapamycin) kinase pathway, which controls translation initiation in yeast⁶, promotes the binding of PP2A to an essential signalling molecule that is known to be conserved in mammals. Thus, the activity of PP2A and other phosphatases may respond to many regulatory inputs.

Camps *et al.*⁴ describe an interaction between the dual-specificity protein phosphatase MKP-3 and its substrate, extracellular-regulated kinase (ERK2), which is a

Table 1 Regulatory interactions involving direct and specific binding of a phosphatase to a kinase

Phosphatase/kinase interaction	Context
Protein serine/threonine phosphatases	
PP2A/Casein kinase-2 α	Mitogen-activated signalling ¹
PP2A/CaMKIV	Ca ²⁺ -induced signalling ³
CaMKIPase/CaMKII	Ca ²⁺ -induced signalling ³
PP2A/Raf kinase	Mitogen-activated signalling ¹⁰
PP1(Reg1)/Snf1	Glucose repression of gene expression ¹¹
Ptc2/Ire1p	Response to unfolded proteins ¹²
Dual-specificity protein phosphatases	
MKP-3/ERK2	Mitogen-activated signalling ⁴
MKP-1/ERK2	Mitogen-activated signalling ¹³
KAP/Cyclin A-Cdk2	Cell-cycle regulation ¹⁴
Protein tyrosine phosphatases	
SHP-1/Src	Signalling in platelets and T cells ¹⁵
PTP1B/p210 bcr-abl	Chronic myelogenous leukaemia ¹⁶
PTP α /p59 ^{lyn}	Cellular signalling ¹⁷

mitogen-activated protein kinase (MAPK). The interaction does not depend on the enzymatic activity of either protein, yet it causes a 30-fold increase in the activity of MKP-3, providing a new example of phosphatase activation through substrate binding. MKP-3 binds specifically to ERK kinases through a non-catalytic region at its amino terminus. Another MAPK phosphatase MKP-4, which has a weakly related amino terminus, binds and is activated by several members of the MAPK family. This implies that the substrate specificity of MKP-3 and other MKPs is conferred by a dedicated amino-terminal kinase-binding domain.

These observations indicate that MAPK phosphatases are highly selective for their binding to particular MAPKs. But, once their selection is made, dephosphorylation occurs very efficiently owing to substrate-triggered activation. MAPK pathways behave like molecular switches, as they flip between the 'on' and 'off' states, in an ultrasensitive response to stimuli that is important for their role in mitogenesis and cell-fate induction⁷. MAPK phosphatases are under the transcriptional control of their substrates, the MAPKs, and this, combined with their substrate-triggered activation, may ensure that the MAPK switch is flicked to the off position as abruptly as it is flicked to the on.

There are several other instances of collaboration between a protein kinase and a protein phosphatase within a self-correcting signalling complex (Table 1). Receptor tyrosine kinases, and receptor tyrosine phosphatases (and their cytosolic cousins) have also been reported to associate within large, multi-protein signalling complexes. Still, there are only a few examples where the identity and regulatory interrelationships of the interacting kinases and phosphatases are well characterized, providing huge scope for future investigation.

The picture that is emerging is one of protein kinases and protein phosphatases operating in concert to provide well-moderated signal transduction that is precise, fast

and, above all, regulated by feedback (Fig. 1). The importance of signalling efficacy is revealed in cancer cells, where the regulation of kinase activity is commonly disrupted by gene mutation or the action of oncogenic viruses. Clearly, where the health of the cell depends on the phosphorylation state of its proteins, what goes up must come down. □

David R. H. Evans and Brian A. Hemmings are at the Friedrich Miescher Institute, PO Box 2543, 4002 Basel, Switzerland.

e-mails: david.evans@fmi.ch

hemmings@fmi.ch

- Hubbard, M. J. & Cohen, P. *Trends Biochem. Sci.* **18**, 172–177 (1993).
- Pawson, T. & Scott, J. D. *Science* **278**, 2075–2080 (1997).
- Westphal, R. S., Anderson, K. A., Means, A. R. & Wadzinski, B. E. *Science* **280**, 1258–1261 (1998).
- Camps, M. *et al. Science* **280**, 1262–1265 (1998).
- Wera, S. & Hemmings, B. A. *Biochem. J.* **311**, 17–29 (1995).
- Di Como, C. J. & Arndt, K. T. *Genes Dev.* **10**, 1904–1916 (1996).
- Ferrell, J. E. Jr & Machleder, E. M. *Science* **280**, 895–898 (1998).
- Heriche, J.-K. *et al. Science* **276**, 952–955 (1997).
- Ishida, A., Kameshita, I. & Fujisawa, H. *J. Biol. Chem.* **273**, 1904–1910 (1998).
- Baccarini, M. (personal communication).
- Ludin, K., Jiang, R. & Carlson, M. *Proc. Natl Acad. Sci. USA* **95**, 6245–6250 (1998).
- Welshinda, A. A., Tirasophon, W., Green, S. R. & Kaufman, R. J. *Mol. Cell. Biol.* **18**, 1967–1977 (1998).
- Sun, H., Charles, C. H., Lau, L. F. & Tonks, N. K. *Cell* **75**, 487–493 (1993).
- Poon, R. Y. C. & Hunter, T. *Science* **270**, 90–93 (1995).
- Somani, A.-K., Bignon, J. S., Mills, G. B., Siminovich, K. A. & Branch, D. R. *J. Biol. Chem.* **272**, 21113–21119 (1997).
- LaMontagne, K. Jr, Flint, A. J., Franza, B. R. Jr, Pendergast, A. M. & Tonks, N. K. *Mol. Cell. Biol.* **18**, 2965–2975 (1998).
- Bhandari, V., Lim, K. L. & Pallen, C. J. *J. Biol. Chem.* **273**, 8691–8698 (1998).

Correction

In the News and Views article "Electronic devices: Single electrons in silicon drops" by D. Christian Glattli (*Nature* **393**, 516–517; 1998), "charging energy" should read "tunnelling energy" (last sentence of paragraph 5); "the first SET to work at room temperature" should be "the first semiconductor SET to work at room temperature (Figure 1 caption); and the year of publication of reference 6 should be 1996.

Daedalus

Extruded tubing

Carbon nanotubes are the most tantalizing of artefacts. They are made by vapour or arc deposition, which gives a very limited product; their unique mechanical and electronic properties can only be glimpsed. Daedalus plans to make them under catalytic control, like any other chain polymer.

Each polymer chain grows from a reactive site on the catalytic surface. When a monomer molecule encounters that site, it reacts to interpose itself between the site and the chain — which is thereupon 'jacked up' by one unit. The process repeats, and a polymer chain grows up from the site.

The monomer of a carbon nanotube is simply the C₂ radical. Daedalus wants to generate these radicals by electrolysis of a melt or solution of carbide ions. Calcium carbide is the obvious candidate: it is made as a melt in an electric furnace. Even so, a solution would be more convenient. Daedalus is hoping to dissolve some alkali metal carbide in a suitable polar but aprotic solvent (water, of course, would decompose it to acetylene).

On electrolysis, such a solution would deposit carbon at the anode. A metal anode would be plated with amorphous carbon or graphite (but, sadly, probably not diamond; a diamond-plating process would be valuable in its own right.) But Daedalus's anode will have the proper catalytic sites to assemble incoming C₂ radicals into carbon nanotubes. It will carry a shape-selective zeolite catalyst comprising circular 'template' sites of the same diameter as the desired nanotubes. With the right organometallic moieties attached to them, they would catalyse the linking of the deposited C₂ radicals into nanotubes.

Thus this cunning anode would grow a 'fur' of carbon nanotubes on its surface. Careful voltage control would be needed, for the tubes are conductors themselves. If the deposition overvoltage were exceeded, they would start to act as anodes in their own right, either lengthening at their far end by direct deposition, or thickening by acquiring successive sleeves of carbon. Longer or thicker tubes could thus be grown on demand.

Nanotube technology will become accessible at last. The promised miracle semiconductors and reinforcing fibres will be here, cheaply and in bulk. Spun and woven into textiles, they might even enter the fashion scene — as run-resistant tights and stockings, or charcoal-grey suiting "Made of real charcoal!"

David Jones

Röntgen's rays

When X-rays were discovered in the last century they swiftly captured the popular imagination, giving rise to a new art form, saucy poetry and circus sideshows alongside their serious roles in science and medicine.

Martin Kemp

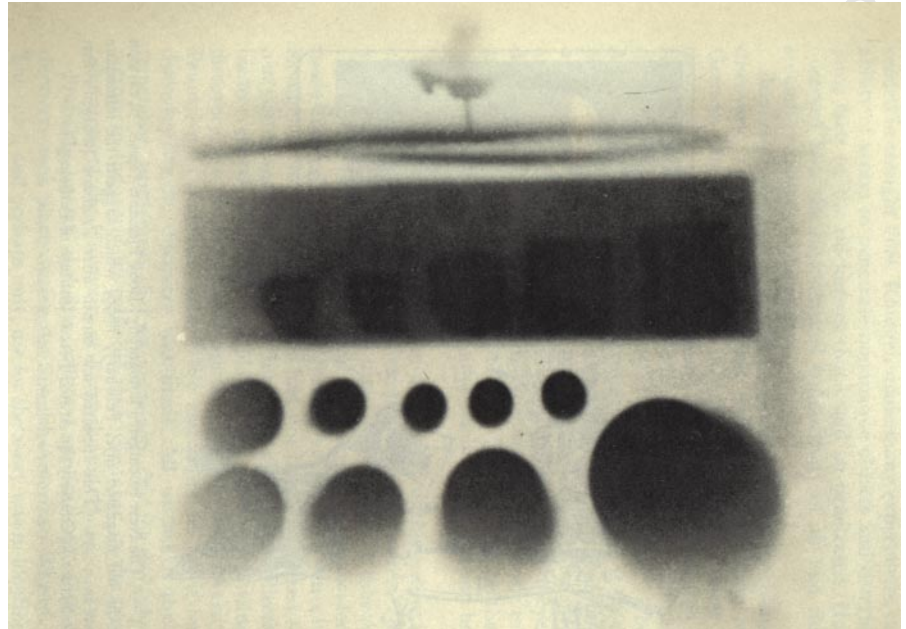
When Wilhelm Röntgen discovered X-rays in 1895, he gave the world a tool that was to revolutionize such sciences as crystallography and medicine. And he also provided for the first time a device that could enable us to 'see' in a way that is untrammelled by the restriction of our visual apparatus to a narrow band in the total spectrum of radiations. The conceptual leap was as great as the technical.

Röntgen discovered X-rays when he was not looking for them. He was pursuing his investigations into cathode rays when he noticed that his Crookes' tube was emitting something that penetrated black paper and travelled further from the tube than forecast. These peculiar emissions not only passed through many solid items placed in their path, including the wooden case of a box of weights, but they also activated phosphorescent materials and imprinted images on photographic plates.

After the main published debut of X-rays in *Nature* on 23 January 1896, the major debates centred not on their uses but on what they were. They were thought of as belonging more in the realm of theoretical research into 'luminiferous ethers' and the nature of electricity than in the world of practical and commercial exploitation — although Thomas Edison characteristically saw that others would grasp "how to profit by it financially".

Alongside the serious science — Röntgen won the first Nobel prize for physics in 1901 — the 'see-through rays' attracted much popular attention, both humorous and alarmist. If we could now see through opaque surfaces, what was there to prevent the making of devices to allow voyeuristic men to inspect the naked bodies of fully clothed women? A poem in the journal *Photography* in 1896 precisely captured the air of circus that inevitably came to surround the discovery, in spite of Röntgen's own public reticence:

**The Roentgen Rays, the Roentgen Rays,
What is this craze:
The town's ablaze
With the new phase
Of X-ray's ways.
I'm full of daze,
Shock and amaze,
For now-a-days
I hear they'll gaze
Thro' cloak and gown — and even stays,
These naughty, naughty Roentgen Rays.**



Röntgen's radiograph of a box of weights.



Josef Maria Eder and Eduard Valenta's "Aesculapian Snake" from *Versuche über photographie mittelst der Röntgenschen Strahlen*, 1896.

The *Pall Mall Gazette* was disgusted by the "revolting indecency of Röntgen rays", and one enterprising London business advertised "X-ray proof underclothing".

The rays also revealed the aesthetic of nature's hidden order. Just one year after Röntgen's announcement, Josef Maria Eder, Viennese photographic chemist and author of authoritative texts on the technical history of photography, and Eduard Valenta, also a photochemist, collaborated on a portfolio of 15 beautifully modulated photogravures of inner skeletal orders.

Like instantaneous photography, X-

radiography simultaneously developed as an important tool of scientific research and as a public performance, exploited across Europe and America by 'scientific' showmen, ignorant of what we now know about the risks of prolonged exposure. As with today's hyped announcements of 'cancer cures', once a discovery leaps out of the professional box, there is little way of controlling its public journeys. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk

Novel retinal photoreceptors

The rod and cone cells of the retina are thought to mediate all of the light responses of the vertebrate eye. But here we show that salmon vertebrate ancient (VA) opsin forms a functional photopigment *in vitro*, and is expressed in a subset of retinal horizontal and amacrine cells *in vivo*. These results indicate that retinal photoreception is not restricted to just rod and cone cells.

Vertebrate photopigments consist of an opsin apoprotein and vitamin A which acts as a chromophore. Opsins can be assigned to one of several well characterized families¹ and are expressed in cells that function as photoreceptors (such as rods, cones, and pinealocytes). Recently, however, complementary DNAs encoding several new opsin families have been identified (peropsin², VA opsin³, parapsin⁴ and melanopsin⁵). Whether these opsins are capable of forming functional photopigments and what role these photopigments could play have been matters of considerable interest⁶.

We analysed the spectral characteristics of *in vitro* generated salmon VA-opsin pigment and its photoproducts by difference spectroscopy. VA photopigment regenerated with 11-*cis*-retinal (vitamin A₁) fits a rhodopsin nomogram with a maximal absorbance at 451 nm (Fig. 1). VA photopigment based upon a vitamin A₂ chromophore would be expected to produce a pigment with a maximal absorbance at 466 nm (ref. 7). The absorbance of photoproducts at 380 nm (Fig. 1) may correspond to either released all-*trans*-retinal or the active state (metarhodopsin II) of VA pigment. We found no light-dependent absorbance changes in sham-transfected COS cell membranes (results not shown). These data demonstrate that, *in vitro*, VA opsin forms a photopigment with a spectral absorbance and photoproducts that are similar to previously described visual pigments.

We determined the sites of VA-opsin expression in the salmon eye using *in situ* hybridization. VA opsin was never observed in the retinal rods or cones, but was restricted to a subset of cells with a location and morphology characteristic of horizontal and amacrine cells (Fig. 2). The numbers of horizontal cells expressing VA opsin varied across the retina (Fig. 2a, b), but were always more than the VA-opsin-expressing amacrine cells. Neither of these cell types has ever been considered to be a photoreceptor. VA opsin was also expressed in cells of the pineal and sub-habenular (data not shown), areas of the brain, which have been implicated in fish photoreception⁸. We are now using electrophysiological methods to determine whether VA-opsin-expressing cells are directly light-sensitive.

What physiological function could VA photoreceptors regulate? In general terms, horizontal and amacrine cells regulate visual receptive fields. Retinal illumination changes their membrane properties, and it has been assumed that all of these effects are mediated by the rods and cones. VA-expressing cells, however, may regulate

receptive fields by a direct light response. The finding of VA opsin in the pineal complex and in the eyes suggests a second function. As both of these areas contain circadian clocks and produce melatonin⁹, VA photoreceptors may mediate the effects of light upon circadian rhythms and/or melatonin synthesis.

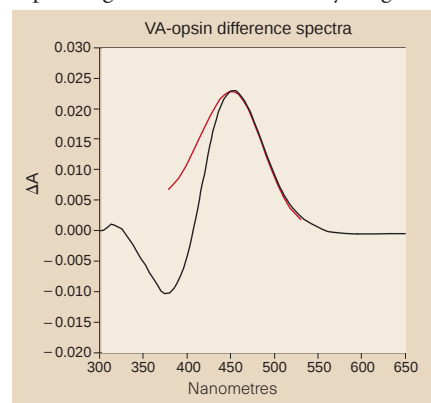


Figure 1 Difference-absorption spectrum of VA opsin. *In vitro* expression and difference spectroscopy have been described¹⁰ and used 50 μ M 11-*cis*-retinal. Light scattering was removed by subtracting a curve fit to data from 550–650 nm. The corrected spectra were smoothed by a running average over 12.5 nm. The rhodopsin nomogram (red) was fitted to the long-wavelength limb of the spectrum as described¹¹.

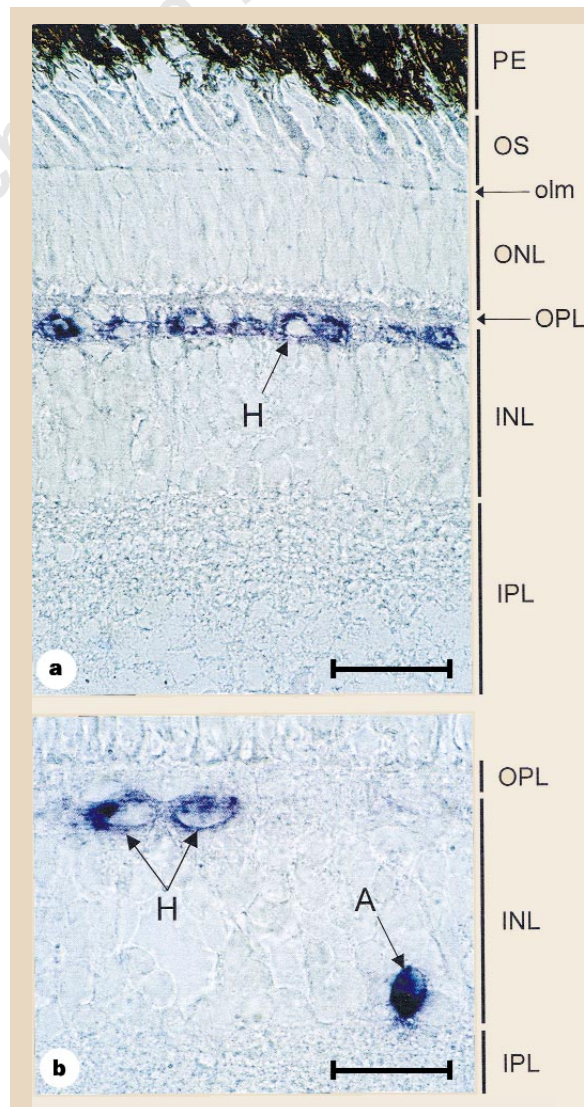


Figure 2 Retinal expression of VA opsin. *In situ* hybridization was done on 10- μ m sections from cryoprotected salmon eyes fixed in 4% paraformaldehyde, using digoxigenin-labelled complementary RNA probes. **a**, Retinal section showing VA-opsin antisense probe labelling. Scale bar, 40 μ m. **b**, Retinal section showing VA-opsin antisense probe labelling in horizontal and amacrine cells. Scale bar, 30 μ m. In neither case did hybridization with sense cRNA probes give any labelling (not shown), confirming the specificity of our procedure. A, amacrine cell; H, horizontal cells; INL, inner nuclear layer; IPL, inner plexiform layer; olm, outer limiting membrane; ONL, outer nuclear layer; OPL, outer plexiform layer; OS, photoreceptor outer segments; PE, pigmented epithelium.

**Bobby G. Soni, Alisdair R. Philp,
Russell G. Foster***

Imperial College of Science, Technology and
Medicine, Department of Biology,
Prince Consort Road, London SW7 2BB, UK

Barry E. Knox

Biochemistry and Molecular Biology,
SUNY HSC at Syracuse, 750 E. Adams Street,
Syracuse, New York 13210, USA

*To whom correspondence should be addressed.

1. Okano, T. *et al.* *Proc. Natl Acad. Sci. USA* **89**, 5932–5936 (1992).
2. Hui, S. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 9893–9898 (1997).
3. Soni, B. G. & Foster, R. G. *FEBS Lett.* **406**, 279–283 (1997).
4. Blackshaw, S. & Snyder, S. J. *Neurosci.* **17**, 8083–8092 (1997).
5. Provencio, I. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
6. Arnheiter, H. *Nature* **391**, 632–633 (1998).
7. Whitmore, A. V. & Bowmaker, J. K. *J. Comp. Physiol. A* **166**, 103–115 (1989).
8. Ekström, P. *et al.* *J. Comp. Neurol.* **265**, 25–33 (1987).
9. Cahill, G. M. & Besharse, J. *Prog. Ret. Eye Res.* **14**, 267–291 (1995).
10. Starace, D. M. & Knox, B. E. *J. Biol. Chem.* **272**, 1095–1100 (1997).
11. Bowmaker, J. K., Astell, S., Hunt, D. M. & Mollon, J. D. *J. Exp. Biol.* **156**, 1–19 (1991).

Do flame retardants threaten ocean life?

Brominated flame retardants are important in modern life. They are used at relatively high concentrations in electronic equipment such as computers and television sets, in textiles, cars and in many other applications. Here we show that two groups of these flame retardants, polybrominated biphenyls (PBBs) and polybrominated diphenyl ethers (PBDEs), are present in sperm whales, which normally stay and feed in deep water, indicating that these compounds have reached deep ocean waters.

The most frequently used brominated flame retardants are tetrabromobisphenol-A, hexabromocyclododecane, PBBs and PBDEs, which are produced globally at an estimated 150,000 tonnes a year^{1–3}. Although

these compounds are similar in behaviour and toxicity to well-known environmental contaminants such as polychlorinated biphenyls (PCBs) and dichlorodiphenyl-trichloroethane (DDT), they have not been banned. Humans may directly absorb PBBs and PBDEs when they are emitted from electronic circuit boards and plastic computer and TV cabinets⁴, and there is also an environmental problem. Because of their high lipophilicity (log K_{ow} > 6, where K_{ow} is the octanol–water partition coefficient) and resistance to degradative processes, PBBs and PBDEs are expected to bioaccumulate easily⁵.

We determined the PBB and PBDE concentrations in 13 marine animals (from 4 species) from coastal seas and the Atlantic Ocean. The sperm and minke whales and whitebeaked dolphin were all stranded alive at the Dutch coast; the juvenile seals were found shortly after they died. We studied the resistance of PBDEs and PBBs to oxidative metabolism by using the cytochrome P450 enzyme system in an *in vitro* bioassay with hepatic microsomes from a whitebeaked dolphin, a sperm whale and a harbour seal. We tested genotoxicity in a mutatox assay.

We found most of the PBBs and PBDEs analysed in both sperm whales and the other samples (Table 1). The total PBDE concentrations in sperm whale blubber (around 100 $\mu\text{g kg}^{-1}$) were about 50-fold higher than the total PBB concentrations in the blubber of the same animals (around 2 $\mu\text{g kg}^{-1}$). The presence of these xenobiotic compounds in sperm whales indicates that the compounds have reached deep ocean waters, as sperm whales are not usually found in shelf seas. Females do not migrate north of a latitude of 45° N (northern Spain); males occur as far north as northern Norway, Iceland and Greenland. At this latitude, sperm whales hunt in waters of depths 400–1,200 m or more.

Postmortem investigations of the same

sperm whales showed a complete absence of food in the alimentary tracts and evidence of weight loss and blubber reduction⁶. Thus it is likely that the PBBs and PBDEs in the blubber and liver of the sperm whales came from deep-water organisms of the oceanic food chain. PCBs are reported to flux at 1.6 $\mu\text{g m}^{-2} \text{yr}^{-1}$ to a depth of 3,200 m in the North Atlantic Ocean (Sargasso Sea)⁷, showing that organic contaminants can be transported to deeper waters fairly quickly.

We found relatively high PBDE concentrations in a whitebeaked dolphin (> 7 mg kg^{-1}) and in harbour seals (> 1 mg kg^{-1}) (Table 1), which had been feeding in the North Sea and the Wadden Sea. PBDE levels in dolphins and seals indicate that ongoing industrial production of PBDEs may create an environmental problem similar to that caused by PCBs, which have been found at concentrations up to 128 mg kg^{-1} in marine mammals⁸.

The *in vitro* biotransformation tests confirm that the two classes of brominated flame retardant are very persistent, even more so than PCBs. Concentrations of BB 15 (4,4'-dibromobiphenyl) decreased by about 80% in the tests, but none of the other PBBs and PBDEs listed in Table 1 showed any indication of biotransformation.

Both classes of compound did not show a genotoxic response in the mutatox assay at concentration ranges of 0.07–900 ng ml^{-1} (PBDEs) and 0.002–9 $\mu\text{g ml}^{-1}$ (PBBs), neither as the parent compound nor after incubation with induced rat liver microsomal fractions ('S9'). However, PBBs can promote the carcinogenicity of other compounds⁹ and PBBs and PBDEs are both listed as compounds that can affect the regulation of thyroid and steroid hormones^{5,10}.

There may be several reasons for the absence of the decabrominated congeners BB 209 and BDE 209 in all wildlife samples. They may disappear (and possibly degrade to lower brominated congeners) because of

Table 1 Concentrations of bromobiphenyls and bromodiphenyl ethers in marine wildlife samples

Species	Tissue	Fat content* (g per kg)	PBB† concentrations ($\mu\text{g per kg wet weight}$)							PBDE† concentrations ($\mu\text{g per kg wet weight}$)			
			15	49	52	101	153	169	209	47	xy‡	99	209
			($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)	($\mu\text{g per kg wet weight}$)
Sperm whale 1	Blubber	722	0.06	0.24	0.40	0.91	1.9	<0.1	<0.5	95	15	26	<6
Sperm whale 2	Blubber	234	0.04	0.13	0.21	0.40	0.73	0.05	<0.3	58	8.1	15	<3
Sperm whale 3	Blubber	317	0.07	0.20	0.36	0.70	1.1	<0.1	<0.4	61	7.5	10	<5
Sperm whale 2	Liver	23	<0.01	<0.01	<0.01	0.63	18	<0.04	<0.3	2.7	0.54	0.91	<3
Whiteb. dolphin	Blubber	990	0.2	7.5	4.1	8.3	13	<0.2	<0.9	5,500	1,200	1,000	<10
Whiteb. dolphin	Liver	27	<0.01	0.06	0.03	0.74	19	<0.02	<0.1	22	5.8	3.0	<1
Minke whale	Blubber	140	0.11	0.27	0.24	0.54	0.82	<0.02	<0.1	88	11	23	<1
Harbour seal 1	Blubber	244	<0.05	34	5.7	9.3	61	12	<1	1,200	110	160	<15
Harbour seal 2	Blubber	963	<0.05	3.1	2.3	1.4	18	<0.2	<1	1,200	100	40	<10
Harbour seal 3	Blubber	722	<0.05	3.0	0.52	1.1	13	<0.1	<1	280	18	140	<10
Harbour seal 2	Liver	35	<0.01	0.10	0.05	0.62	1.5	<0.02	<0.1	21	0.93	0.85	<2
Harbour seal 3	Liver	51	<0.01	0.10	0.03	0.04	0.82	<0.01	<0.1	12	0.33	5.1	<1
Harbour seal 4	Liver	30	<0.01	0.90	0.14	0.44	13	<0.02	<0.1	20	0.07	0.53	<2
Mackerel	Muscle	152	0.01	0.01	0.01	<0.01	0.04	<0.03	<0.2	5.4	1.8	1.9	<2

* Total lipid contents are according to ref. 11, except for seal 1 blubber and seal 3 blubber and liver, for which extractable lipid contents are shown.

† Numbering of PBB and PBDE congeners is according to the PCB numbering system of ref. 12.

‡ xy-PBDE is pentabrominated BDE with unknown structure.

chemical or microbial degradation (weathering), or they may be taken up at reduced rates because it is difficult for their relatively large molecules to cross cell membranes. More research is required to understand the environmental behaviour of the decabrominated compounds, which, despite being increasingly produced, have not yet been found in aquatic organisms.

The presence of PBBs and PBDEs in sperm whales, the high levels of particularly PBDEs in seals and dolphins, and the ongoing industrial production of these compounds suggest that an environmental problem may be on its way.

Jacob de Boer, Peter G. Wester

DLO-Netherlands Institute for Fisheries Research, PO Box 68, 1970 AB IJmuiden, The Netherlands
e-mail: j.deboer@rivo.dlo.nl

Hans J. C. Klamer

National Institute of Coastal and Marine Management/RIKZ, PO Box 207, 9750 AE Haren, The Netherlands

Wilma E. Lewis, Jan P. Boon

Netherlands Institute for Sea Research (NIOZ), PO Box 59, 1790 AB Den Burg, Texel, The Netherlands

- Sellström, U. *Polybrominated Diphenylethers in the Swedish Environment* (ITM report 1996, 45, Stockholm Univ., 1996).
- Zitko, V. *Mar. Pollut. Bull.* **26**, 584–585 (1993).
- Sellström, U. & Jansson, B. *Chemosphere* **31**, 3085–3092 (1995).
- Zelinski, V. *et al. Chemosphere* **27**, 1519 (1993).
- Brominated Diphenylethers* (IPCS, Environmental Health Criteria no. 162, WHO, Geneva, 1994).
- Jauniaux, T. *et al. J. Wildlife Dis.* **34**, 99–109 (1998).
- Knap, A. H., Binkley, K. S. & Deuser, W. G. *Nature* **319**, 572–574 (1986).
- Boon, J. P. *et al. Arch. Contam. Toxicol.* **33**, 298–311 (1997).
- Silberhorn, E. M., Glauret, H. P. & Robertson, L. W. *Crit. Rev. Toxicol.* **20**, 440–496 (1990).
- Polybrominated Biphenyls* (IPCS, Environmental Health Criteria no. 152, WHO, Geneva, 1994).
- Blythe, E. G. & Dyer, W. J. *Biochem. Physiol.* **37**, 911–917 (1959).
- Ballschmiter, K. *et al. J. High Res. Chromatogr.* **15**, 260–270 (1992).

Density of states reflects diameter in nanotubes

Dresselhaus¹ asked recently why scanning tunnelling microscopy and spectroscopy experiments^{2,3} find a similar number of peaks, in the density of states near the Fermi level, for a range of semiconducting single-wall carbon nanotubes with roughly the same diameters. After all, semiconducting achiral zigzag single-wall nanotubes can have translational unit cells more than an order of magnitude smaller than those of chiral semiconducting single-wall nanotubes with a similar diameter, so they might be expected to have fewer but sharper peaks in their density of states¹. We show that all semiconducting single-wall nanotubes with similar diameters should have a similar density of states near the Fermi level, independent of translational unit cell

Figure 1 Density of states in nanotubes. **a**, Hexagonal central Brillouin zone of graphene. Parallel lines depict allowed states for (11,9) SWNT. Circle at bottom right encloses region of states near ϵ_F . **b**, Expanded depiction of allowed states near ϵ_F , with dotted line parallel to $\mathbf{R}$, k_F corner of hexagon with energy ϵ_F , and k_A , k_B and k_C special points of closest approach to k_F of three nearest allowed-state lines in reciprocal space. **c**, DOS per carbon, $\rho(E)$, as a function of energy in units of $\alpha = |V_{pp\pi}|d/D$ computed using the full graphene dispersion relation $\epsilon_{\pm}(\mathbf{k}) = \pm |V_{pp\pi}|[3 + 2\cos(\mathbf{k} \cdot \mathbf{R}_1) + 2\cos(\mathbf{k} \cdot \mathbf{R}_2) + 2\cos(\mathbf{k} \cdot (\mathbf{R}_1 - \mathbf{R}_2))]^{1/2}$

for six different SWNTs with similar diameters, D , for a range of chiral angles, ϕ , and number of atoms in the translation unit cell, N_T . The average D for these six tubes is $\bar{D} = 1.33$ nm for $d = 0.142$ nm with all D varying less than 3% from $\bar{D}$. Also, $\alpha = 0.29$ eV for $V_{pp\pi} = -2.7$ eV and $D = \bar{D}$. All DOS are broadened by a lorentzian of width 0.01α . Shown are the DOS for the (17,0) [solid line; $D = 1.33$ nm, $\phi = 0^\circ$, $N_T = 68$], (13,6) [dashed line; $D = 1.32$ nm, $\phi = 18^\circ$, $N_T = 1,132$] and (11,9) [dotted line; $D = 1.36$ nm, $\phi = 27^\circ$, $N_T = 1,204$] semiconducting SWNTs. Vertical lines A, B and C (corresponding to k_A , k_B , and k_C in Fig. 1b) are peak positions in the occupied semiconducting DOS predicted using $\epsilon_{-}(\mathbf{k}) = -\frac{3}{2}|V_{pp\pi}|d/|\mathbf{k} - \mathbf{k}_F|$ approximately valid for $|\epsilon_{-} - \epsilon_F| \leq \frac{3}{2}|V_{pp\pi}|$. Also shown are the DOS for the (10,10) [solid line; $D = 1.36$ nm, $\phi = 30^\circ$, $N_T = 40$], (14,5) [dashed line; $D = 1.34$ nm, $\phi = 15^\circ$, $N_T = 1,164$] and (16,1) [dotted line; $D = 1.29$ nm, $\phi = 3^\circ$, $N_T = 1,092$] metallic SWNTs offset upward by 0.03. Vertical line A' is the first peak position in the occupied metallic DOS predicted from the approximate theory. Curvature introduces a small gap at ϵ_F in the non-armchair metallic tubes which is probably smeared out in the experimental results^{2,3}.

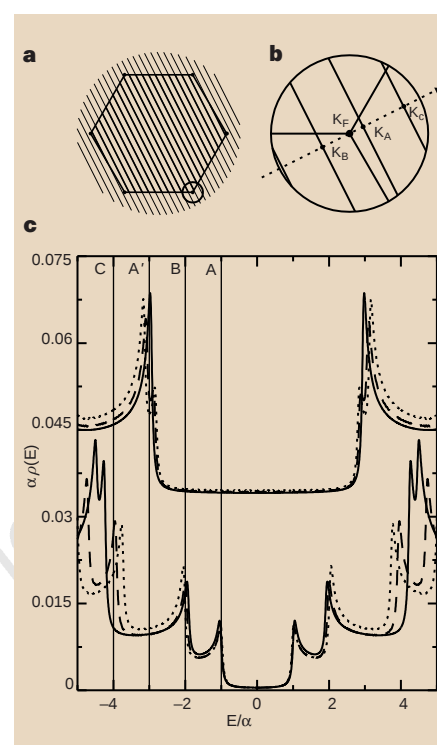
size or chiral angle.

A single-wall nanotube (SWNT) can be constructed by rolling up a single sheet of graphite (graphene) along one of its two-dimensional lattice vectors $\mathbf{R} = n_1\mathbf{R}_1 + n_2\mathbf{R}_2$ to form an (n_1, n_2) nanotube with diameter $D = |\mathbf{R}|/\pi$. To a first approximation, the allowed electron states of the SWNT are the graphene bands satisfying the periodic boundary conditions $\mathbf{k} \cdot \mathbf{R} = 2\pi m$, with m being an integer.

Graphically, these allowed $\mathbf{k}$ values lie along the parallel lines in Fig. 1a, which are perpendicular to $\mathbf{R}$ with an interline spacing of $2/D$. The Fermi level, ϵ_F , is given by states at the corner points of the hexagonal Brillouin zone in Fig. 1a. So, neglecting the effects of curvature, nanotubes are either metallic or semiconducting depending on whether or not $\mathbf{k}_F \cdot \mathbf{R} = 2\pi m$, where $\mathbf{k}_F = 4\pi(\mathbf{R}_1 - \mathbf{R}_2)/9d^2$ and d is the carbon-carbon bond distance⁴.

The two-dimensional dispersion relation of the occupied states of graphene is given to lowest order in $d|\mathbf{k} - \mathbf{k}_F|$ by $\epsilon_{-}(\mathbf{k}) = -\frac{3}{2}|V_{pp\pi}|d|\mathbf{k} - \mathbf{k}_F|$, where $V_{pp\pi}$ is the nearest-neighbour $pp\pi$ interaction⁵. Hence, near ϵ_F , $\epsilon_{-}(\mathbf{k})$ is radially symmetric about the point k_F located by $\mathbf{k}_F$ and decreases with increasing $|\mathbf{k} - \mathbf{k}_F|$.

For semiconducting SWNTs, the point of closest approach to k_F in any allowed line therefore yields to good approximation a local maximum in the one-dimensional



band structure as long as $d|\mathbf{k} - \mathbf{k}_F| \ll 1$, leading to a van Hove singularity and a divergence in the occupied density of states (DOS) near ϵ_F . These special points also yield peaks in the unoccupied DOS because the dispersion relation for the unoccupied states is given by $\epsilon_{+}(\mathbf{k}) = -\epsilon_{-}(\mathbf{k})$. The special point on the line closest to k_F (labelled as k_A in Fig. 1b) also yields the SWNT band gap^{5,6}. This point is separated from k_F by a distance $|\mathbf{k}_A - \mathbf{k}_F| = \frac{2}{3D}$, one-third the interline spacing — a result easily confirmed by projecting $\mathbf{k}_A - \mathbf{k}_F$ along $\mathbf{R}$. So k_A generates peaks in the DOS near $\epsilon_{\pm}(\mathbf{k}_A) = \pm |V_{pp\pi}|d/D$ and a band gap near $2|V_{pp\pi}|d/D$ when $\frac{2d}{3D} \ll 1$.

Incrementing by the interline spacing, $2/D$, the next two special points yield peaks in the DOS near $\epsilon_{\pm}(\mathbf{k}_B) = 2\epsilon_{\pm}(\mathbf{k}_A) = \pm 2|V_{pp\pi}|d/D$ when $\frac{4d}{3D} \ll 1$, and peaks near $\epsilon_{\pm}(\mathbf{k}_C) = 4\epsilon_{\pm}(\mathbf{k}_A) = \pm 4|V_{pp\pi}|d/D$ when $\frac{8d}{3D} \ll 1$. Changing chirality at a given SWNT diameter effectively rotates the lines of allowed states around k_F without changing the distances between the points along the lines and k_F . Therefore the DOS of two semiconducting SWNTs with differing chiral angles but similar diameters are similar in the vicinity of ϵ_F because $\epsilon_{\pm}(\mathbf{k})$ is radially symmetric about k_F . A parallel analysis shows that all metallic SWNTs with similar diameters have similar DOS near ϵ_F within the graphene sheet model but with peaks close

to $\pm 3|V_{pp\pi}|d/D$ when $\frac{2d}{D} \ll 1$.

All of these predictions are consistent with the numerical results shown in Fig. 1c. Because the semiconducting and metallic SWNTs studied experimentally all have $d/D \approx 0.1$ and $V_{pp\pi} \approx -2.7$ eV, these predictions explain the observed peak spacing in the SWNT DOS^{2,3}, further confirming the band structure theory of these nanowires⁴⁻⁹.

C. T. White, J. W. Mintmire

Code 6179, Naval Research Laboratory,
Washington, DC 20375, USA
e-mail: white@alchemy.nrl.navy.mil

1. Dresselhaus, M. S. *Nature* **391**, 19–20 (1998).
2. Wildöer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. *Nature* **391**, 59–62 (1998).
3. Odom, T. W., Huang, J.-L., Kim, P. & Lieber, C. M. *Nature* **391**, 62–64 (1998).
4. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* (Academic, New York, 1997).
5. Mintmire, J. W., Robertson, D. H. & White, C. T. *J. Chem. Phys. Solids* **54**, 1835–1840 (1993).
6. White, C. T., Robertson, D. H. & Mintmire, J. W. *Phys. Rev. B* **47**, 5485–5488 (1993).
7. Mintmire, J. W., Dunlap, B. I. & White, C. T. *Phys. Rev. Lett.* **68**, 631–634 (1992).
8. Hamada, N., Sawada, A. & Oshiyama, A. *Phys. Rev. Lett.* **68**, 1579–1581 (1992).
9. Saito, R., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. *Appl. Phys. Lett.* **60**, 2204–2206 (1992).

A more reliable design for biodiversity study?

Naeem and Li¹ present the results of a microcosm study in which species diversity of organisms within trophic groups was varied. They conclude that the existence of multiple species within these groups enhanced the “reliability” of these systems, that is, the increased likelihood of a consistent level of performance over a given unit of time. But there are problems with their study. For the least diverse communities, one predator species was randomly selected from a selection of two, one autotroph from a selection of three, one consumer of bacteria from a selection of five, and one omnivore from a selection of six. Meanwhile, with the most diverse communities, both predator species and all three autotroph species were used, three consumer bacteria were chosen from the five and three omnivores from the six.

Thus, for the most diverse communities, the species composition of the individual

replicates was identical for three of the five functional groups considered (when the non-manipulated functional group, decomposers, is also considered) and, for the least diverse communities, the composition was identical across replicates for only one of the functional groups. Even for the remaining two groups, there is a greater probability of the species compositions being more similar to each other across replicates in the more diverse treatment (Table 1).

Therefore, the observation that biomass variation of the non-manipulated functional group (decomposers or bacteria) across replicates was less in the more diverse microcosms (used by Naeem and Li as evidence that enhanced diversity increased “ecosystem reliability”) could be explained simply in terms of the individual replicates of the more species-rich microcosms being more similar to each other than those of the less diverse microcosms at the start of the experiment, in terms of both the composition of food sources and the consumers of bacteria.

Similarly, the observation that algal density was most closely associated with the number of functional (trophic) groups in the more species-rich microcosms could be attributed to lower variability in starting conditions across replicates of these microcosms, resulting in a tighter and more consistent relationship across replicates within treatments that is less likely to be obscured by within-treatment variation.

The hypothesis tested by Naeem and Li could only be investigated reliably by using a sufficiently large species pool to enable non-overlapping species compositions across replicates of the most species-diverse microcosms. Such an experiment could consist of, for example, four different groups (with non-overlapping compositions) of ten sets of identical high-diversity communities and four different groups of ten sets of identical low-diversity communities.

The variability is then calculated for each group of ten, and the variability across each group of ten is used as one datum. The result is four replicate measures of variability for each of the high- and low-diversity situations. But even with a larger species pool, caution is needed in the interpretation of results when the causative effects of biodiversity are being investigated²⁻⁵. A more robust experimental design is needed to justify claims such as “provision of

adequate redundancy may be one reason for preserving biodiversity”.

David A. Wardle

Landcare Research, PO Box 69, Lincoln 8152,
New Zealand
e-mail: wardled@landcare.cri.nz

1. Naeem, S. & Li, S. *Nature* **390**, 507–509 (1997).
2. Grime, J. P. *Science* **277**, 1260–1261 (1997).
3. Huston, M. A. *Oecologia* **110**, 449–460 (1997).
4. Wardle, D. A., Zackrisson, O., Hörnberg, G. & Gallet, C. *Science* **278**, 1867–1869 (1997).
5. Tilman, D. et al. *Science* **278**, 1866–1867 (1997).

Naeem and Li reply — Wardle’s concerns arise from conflating biodiversity loss within an ecosystem (our study¹) with biodiversity loss among ecosystems. To examine biodiversity loss within an ecosystem experimentally, a community is compared with more depauperate versions of itself^{2,3}. This design mimics the pattern observed in nature and sheds light on theory and mechanisms. Our experiment followed this approach to test predictions of reliability theory based on the mechanism of compensatory growth among redundant species.

In contrast, to examine biodiversity loss among ecosystems, communities of similar biodiversity but with unique species composition are compared across different diversity levels. (So far, only the European Union’s recent BIODEPTH experiment follows such a design.) So what Wardle calls a “more robust” design would actually be addressing a different question.

Wardle’s concerns arise because he believes that variability in initial community composition will be correlated with variability in final measures of ecosystem functioning, but he does not provide any theory or mechanism for this hypothetical correlation. Trivially, communities of non-overlapping composition may show less variability within communities than among them (Wardle’s design) but there is no ecological foundation for the belief that variability within replicates of a diverse community will be lower than variability among depauperate versions of itself (our design).

Factors such as local extinction, variation in nutrients or energy output, and increases in density and standing biomass by several orders of magnitude over many generations (all characteristics of replicate communities in our experiment) generate a wide variety of production and dynamic responses in communities. Contrary to Wardle’s claim, our findings clearly support the possibility that preserving an ecosystem’s biodiversity might provide more reliable functioning.

Shahid Naeem, Shibin Li

Department of Ecology, Evolution and Behavior,
University of Minnesota, St Paul,
Minnesota 55108, USA
e-mail: naeem001@maroon.tc.umn.edu

1. Naeem, S. & Li, S. *Nature* **390**, 507–509 (1997).
2. Naeem, S. et al. *Nature* **368**, 734–737 (1994).
3. Tilman, D. et al. *Science* **277**, 1300–1302 (1997).

Table 1 Expected proportion of total number of species in common between any two replicates for each diversity treatment at the start of the experiment conducted by Naeem and Li

Functional group	Lowest diversity treatment	Intermediate diversity treatment	Highest diversity treatment
Autotrophs	0.333	0.667	1.000
Decomposers	1.000	1.000	1.000
Omnivores	0.167	0.333	0.567
Consumers of decomposers	0.200	0.400	0.667
Top predators	0.500	1.000	1.000
All species	0.368	0.579	0.776

Selling science against the odds

Science in Public: Communication, Culture and Credibility

by Jane Gregory and Steve Miller

Plenum: 1998. Pp. 282. \$29.95, £18.15

Communicating Research

by A. J. Meadows

Academic: 1998. Pp. 224. \$59.95, £39.95

Graham Farmelo

Hard though it is to believe, as one sees one's colleagues sneak off early to watch the World Cup, Britain's national sport is no longer soccer, but the Lottery. Twice a week, millions of Britons spend their money on tickets, even though they stand less chance of winning than they have of being struck by a comet. And if such a collision were imminent, surveys tell us that only about 48 per cent of UK adults would be most interested in hearing the views of scientists, rather than other professionals, and that almost half as many would prefer to listen to an astrologer.

All this is far from encouraging for those of us who spend most of our lives communicating science to the public. Over the past two decades in the West, substantial funds and inordinate amounts of scientists' time have been lavished on the promotion of what is widely and unsatisfactorily known as the public understanding of science, or scientific literacy. We practitioners are now even said to be part of a 'movement', heaven help us.

What impact have we made? In 1995, the veteran science educator Morris Shamos wrote a scathing critique of the entire public-understanding project in his underrated book *The Myth of Scientific Literacy* (Rutgers University Press). Having analysed the apparently feeble results of decades of public-understanding initiatives in the United States, Shamos concluded that the hope of achieving a universal, basic level of scientific literacy is "little more than a romantic idea". Much of his analysis applied equally well to other western countries, especially the United Kingdom. So did his book inspire a widespread re-evaluation of policy? Not at all: the movement was sublimely unperturbed.

But Shamos's analysis is taken seriously in Jane Gregory and Steve Miller's concise review of the history, theory and contemporary practice of the public understanding of science. To their credit, the authors present the views of Shamos and other dissenters, as well as the opinions of a wide range of experts, including sociologists, philosophers, historians, media experts and scientists.

Over the past 15 years or so, the aims of practical public-understanding initiatives have changed. In the 1980s there was an emphasis on what science the ignorant pub-

lic needs to know (the 'deficit model'), but this has been superseded by a concern to facilitate dialogue between scientists and lay people to encourage mutual trust. Through a series of well-chosen case studies, Gregory and Miller persuade us that the public have every reason to doubt blithe Orwellian reassurances that it would be in their best interests if they kept conveniently quiet and paid science the usual civil inattention.

Whether the public will ever come to terms with risk assessment is a moot point. It is difficult enough to discuss risk from the statistical and psychological points of view when the precise nature of the risk is known, but when it is not, scientists and governments are in a difficult position. If they assume the worst they are accused of scare-mongering; if they lean towards optimism they are charged with negligence.

Gregory and Miller handle this topic very effectively by considering three examples: the risk of a global catastrophe following a collision between the Earth and a comet; that of getting cancer from eating apples sprayed with the ripening agent Alar; and that of getting Creutzfeldt-Jakob disease (CJD) from eating beef infected with bovine spongiform encephalopathy (BSE, or 'mad cow disease'). Their review of the continuing CJD tragedy is sensitive and level-headed,

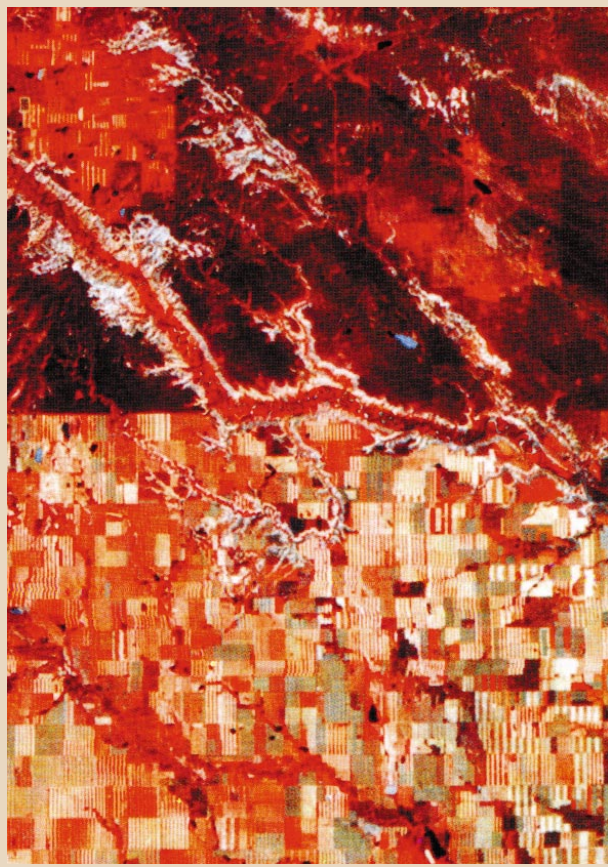
but it is a pity that they concentrate so strongly on the UK perspective. They fail to mention the US meat industry's case against the talk-show host Oprah Winfrey, who said, following the CJD scare, that she intended never to eat another hamburger. For many US readers, the omission of this part of the story will exemplify the authors' apparent Britocentricity.

Another example of this occurs in their treatment of museums, to which the authors generously devote one of the book's longest chapters. Gregory and Miller are as thorough as ever, although their account of the impact that exhibitions and events in museums and science centres can make on their visitors is surprisingly superficial. They seem to be unaware that most of the important work in exhibition evaluation has been done in the United States; almost all of the most influential references are missing.

But it would be mean to complain too much. Gregory and Miller have produced a splendidly readable and thorough review of their field. The even-handedness, comprehensiveness and sheer accessibility make it an indispensable text for students of the public understanding of science. It is also a valuable read for scientists who are now, more than ever, under pressure to talk with the populace rather than at them.

Taking the long view

Remote sensing dates back to the aerial photographs used for military intelligence in the American Civil War (1861–65), but it was the images returned by the US space programme a century later that first captivated public attention. In *Images of the Earth: A Guide to Remote Sensing*, 2nd edn (Oxford University Press, £25, \$45), S. A. Drury uses a wealth of images to illustrate the importance of remote sensing in addressing environmental and social problems "on the scale at which they really present themselves". The Landsat false-colour image on the right highlights the contrast in land use across a 70-km stretch of border between the United States and Canada.



The traditional attitude of scientists towards science was well summarized in the 1940s by the sociologist Robert Merton, who spoke of the imperative of sharing information and of science's characteristic disinterestedness, universalism and organized scepticism. Whereas Gregory and Miller report the attacks on this idealistic picture made by Bruno Latour and other influential scholars, Jack Meadows scarcely mentions them in his new book.

Meadows discusses how academic researchers in the sciences communicate their findings to each other and to the public. This is scientific research as many scientists would have it portrayed, unpolluted (as they would say) by the subversive misreadings of the so-called sociologists of scientific knowledge. No wonder the book is endorsed by Lewis Wolpert, the sociologists' scourge, who commends it as "sociology of science at its best".

Most scientists will find Meadows' account predictable. Many, I fear, will see it simply as the elevation of their informal professional chatter to the level of a worthy academic discourse. Is it easier to publish in the sciences than in the humanities? When do most scientists do their most intensive reading of the journals? Questions such as these are answered authoritatively here, with the aid of the best available data, providing answers that usually confirm the prejudices of the seasoned coffee-table pundit.

Meadows' approach may be unfashionable but he makes many points that ring a good deal truer than the writings of some of the more revisionist sociologists of science. For example, he recounts how the late Lev Landau rated his fellow physicists on a logarithmic scale: he put Einstein in class 0.5, Dirac and Heisenberg in class 1, himself in class 2.5, with most humble foot soldiers of physics way down below. Hard to defend though this exercise is, its results chime better with the experience of many physicists than observations of the unimportance of personalities made by some of the modish critics of science, many of whom do not seem to have done any science.

A meeting of minds between scientists and the scholars who study their work remains a distant hope. Similarly, Gregory and Miller point out that there is a worrying chasm in the field of public understanding of science between its theorists and its practitioners, who work with the public and the media. If a harmony could somehow be achieved, scientists might stand a better chance of winning back public trust, now at a worryingly low level throughout the West. You never know, we might live to see a day when astronomers outnumber astrologers, if not footballers. □

Graham Farmelo is at the Science Museum, South Kensington, London SW7 2DD, UK.

Tripped up by timekeeping

Time's Pendulum: The Quest to Capture Time – From Sundials to Atomic Clocks

by Jo Ellen Barnett

Plenum: 1998. Pp. 340. \$27.95, £16.94

Kristen Lippincott

The only problem with the recent flowering of 'popular science writing' is that the large number of authors who have succeeded in creating exciting, readable and intelligent books has led a series of other writers to think that writing accessible science books must be easy. Unfortunately, on the back of the wave, we are now being forced to suffer the consequences. Those who are not demonstrably well read, nor expert in any given field, nor (in many cases) particularly gifted communicators, have decided that now is the time to enter the fray.

Time is a tricky subject to write about clearly. When trying to define time, even the usually eloquent St Augustine became tongue-tied. In trying to write a popular book about the ways in which Western man has developed a timekeeping system, however, it is a great mistake to assume that the

model of 'popular science writing' is going to help. The story of time (insofar as it might have one) is not a story of discovery or progress. There may be heroes in the story of time measurement or timekeeping, but this — as all my horologically inclined colleagues are eager to tell me — has very little to do with time itself. Indeed, even at the end of the twentieth century, our understanding and modes of thinking about time are more closely linked to cultural assumptions, religion and philosophy than to what most *Nature* readers would understand as science.

If one accepts that there might be intent behind the subtitle of Jo Ellen Barnett's book, one should also recognize it as a warning. The claim that this book explains "the quest to capture time — from the sundial to the atomic clock" presupposes two things: first, that there has been a quest to capture time; and second, that there is some sort of triumphal progression from a timekeeper that marks the apparent passage of the Sun across the sky to a timekeeper that measures the rates of oscillation in an atom of caesium-133. Both are claims that, on reflection, are difficult to sustain; but such is Barnett's starting point.

The rest of the book is a series of largely unrelated anecdotes strung together to show how silly ancient man was and how clever modern man is. To wit: "Our ancestors, of



The times they are a-changin': a Byzantine sundial calendar and a caesium atomic clock.

course, had no idea of the things discussed in the previous chapter. They were at least as ignorant of our twenty-four *equal* [her italics] hours... as they were about any of our astronomical discoveries"; or "The more complex life styles they [the ancient Egyptians] bred would certainly have been made easier by a firmer awareness of the time of day"; or "So powerful was the belief that time somehow inhered in [sic] the rhythms of nature that it had kept timekeeping in bondage to our planet's cycle of light and darkness for over three millennia".

Moreover, this positivistic romp through a self-styled history of time results in some fairly stunning anachronisms. When Steno decided to become ordained as a Roman Catholic priest, it was because he "couldn't handle the discontinuity with the biblical time line already hinted at by his discoveries"; and "Darwin himself may have even been behind the times when he set out on the *Beagle* in 1831... [because] he believed in 'the strict and literal truth of every word in the Bible'".

Even if one were willing to suspend one's better judgement and try to enter into the spirit of the argument, the text is so peppered with errors, misunderstandings and sentences whose meaning remains a complete mystery that moving from page to page becomes a bit of a chore. To list even a fraction of the mistakes and inanities would be a waste of the reader's time.

As time is indeed a tricky subject, I tend to give the benefit of the doubt to anyone who even attempts to tackle it. But this book drove me well beyond incredulity and perilously close to despair. The only advice to offer is: for an intelligent discourse on time, go and re-read G. J. Whitrow, Derek Howse, David Landes, Jacques LeGoff, Gerhard Dohrn-van Rossum, Arno Borst or even Mircea Eliade — but steer well clear of this book. □

Kristen Lippincott is at the Old Royal Observatory, Greenwich, London SE10, UK.

Truth and consequences

A Tremor in the Blood: Uses and Abuses of the Lie Detector, 2nd edition

by David T. Lykken

Plenum: 1998. Pp. 333. \$28.95, £17.55

Gisli H. Gudjonsson

Lying and its detection hold a fascination for many people. The objective measurement of physiological responses such as heart rate, blood pressure and electrodermal reactivity forms the basis of lie detection. It is therefore not surprising that machines developed for measuring such responses in a medical set-



The nose has it: the specific physiological response to lying that has eluded modern science.

ting are of interest to researchers, police officers and practitioners with an interest in lie detection.

David T. Lykken provides a fascinating insight into the common abuses of this potentially important scientific tool. His main message is that lie detection techniques, apart from his own 'guilty knowledge technique', of course, are fundamentally flawed, prone to high error rates, and every year result in several miscarriages of justice. This is a big blow to the advocates of the most common techniques, such as the 'control question test'. The book is written by an eminent scientist for non-scientists, including lawyers, policy-makers and those who feel aggrieved by the injustice of having been wrongly classified as deceptive after allegedly failing a test. The first edition of the book appeared in 1981 (McGraw-Hill).

Lykken has for the past 20 years been the most vocal authority opposing the use of the polygraph lie detector. His interest in lie detection dates back to experiments he conducted in the late 1950s and his pioneering

work into the use of the 'guilty knowledge technique'. He is also well known and respected for his work on the genetic features of twins.

He tells us that the first instrument to detect lies was used at the end of the nineteenth century, when the Italian criminologist Cesare Lombroso used a 'plethysmograph' to measure changes in blood pressure as a means of detecting deception. But it was not until the early twentieth century that William Marston, a student of Hugo Munsterberg at Harvard, invented the lie detector and claimed to have discovered a 'specific lie response', which later proved to be unfounded. Indeed, a specific physiological response to lying has never been found, and probably never will be.

In 1915, Marston argued that lying was accompanied by an increase in systolic blood pressure, and about 20 years later he described experiments that showed a successful detection rate of between 97 and 99 per cent. Similarly extravagant claims were made by other early polygraph researchers,

including the Reverend Walter Summers, who, in 1937, claimed 100 per cent accuracy in detecting lying in 200 criminal cases by the use of an electrodermal apparatus.

Lykken points out that similar extravagant claims are also stated by many contemporary workers, who claim the lie detector to be at least 98 per cent effective in detecting deception. Lykken explains such unrealistic claims, which contrast greatly with the objective and scientifically based evidence, by the fact that polygraph examiners rarely know the real truth about the subject's guilt or innocence and, when a deceptive outcome is confirmed by a subsequent confession, this reinforces their faith in the efficacy of the instrument and their own interpretations of the polygraph charts.

The apparently blind faith that many polygraph examiners have in their own test results is also shared by many of those who agree to take the test without realizing its limitations and the potential personal consequences. Failing a polygraph test, Lykken cogently argues, often gives employers an excuse to dismiss staff, and in criminal cases it sometimes results in wrongful convictions.

The 'guilty knowledge technique', or 'concealed information test' as David Raskin, Lykken's main opponent, prefers to call it, is an indirect test of deception. The purpose of the technique is to detect concealed information rather than lying *per se*. Lykken, not surprisingly, advocates the use of his technique and describes its advantages over other techniques, including its sound theoretical base. Given the technique's apparent superiority, Lykken appears justifiably surprised by the resistance of polygraph examiners to applying his technique in criminal investigations. There are, of course, often great difficulties in using the test in practice because of the absence of special knowledge items that make up the test; this he fully acknowledges. Sadly, Lykken's technique, though important, has never really caught on.

Lykken illustrates the role of the polygraph in eliciting confessions from those who fail the test. Many polygraph examiners appear to feel justified in coercing a confession out of a suspect who may be innocent of the crime of which he or she is accused. Amazingly, up to 75 per cent of this group make a confession. How many are false? There are certainly reported cases of polygraph examiners falsely persuading people of their guilt. It is a potentially dangerous tool indeed.

Finally, Lykken discusses the dangers of using polygraph test results in court. The main problem is that the apparent scientific status of the polygraph, and people's blind faith in the instrument, often proves prejudicial to the defendant. Polygraph techniques appear to offer so much but often deliver so little. They should be used cautiously and

selectively. Lykken's presentation of the evidence, both for and against the polygraph, may seem to some to be one-sided and biased, and at times he comes across like an advocate rather than an objective scientist. But he has made a strong case, which should be taken seriously. □

Gisli H. Gudjonsson is at the Institute of Psychiatry, King's College, University of London, London SE5 8AF, UK.

Digging up history's skeletons

Bones in the Basement: Postmortem Racism in Nineteenth-Century Medical Training

by Robert L. Blakely and Judith M. Harrington

Smithsonian Institution Press: 1998. Pp. 380. \$45, £34.95

Ruth Richardson

The old building of the Medical College of Georgia, Augusta, is now a US national historic landmark. In the nineteenth century it housed the college dissecting rooms, so it was perhaps not surprising that bones were discovered in its basement when work to transform it into a museum began in 1989.

Robert L. Blakely and students from his class in forensic anthropology were summoned to undertake 'salvage archaeology'. They dug three sample areas and, together with the subsequent efforts of the builders, unearthed 2,000 artefacts, 300 animal bones and more than 9,000 human bones and bone fragments. The site, described as a 'medical dump', is about a metre deep throughout, with "accretional layering [and] jumbled deposits". No stratigraphy was undertaken, so it is not known whether the datable artefacts were laid down in chronological order, or whether the bones buried with them are contemporary. Window glass from 1840 may have remained intact until 1900.

The dig's haste perhaps excuses some of its shortcomings, but not those of the book, which has been nearly ten years in the making. It is poorly organized. The building ceased to be a medical school in 1912, but a discussion of modern dissection appears before the reader is introduced to nineteenth-century medical practices, such as grave robbery. Chapter 3 — on medicine bottles found in the basement — was written in ignorance, it seems, of the finding in chapter 11 that African-Americans used them as burial-site adornments. One reaches chapter 10 before being vouchsafed insights into the profound problems of palaeopathology, told some of the reasoning behind the book, and informed of key elements of the research methodology. For example, researchers

worked independently to minimize intra-observer error and bias, and bones were analysed as components of a population rather than as individual finds.

These points explain the existence of contradictory findings in the book, but not why the volume leaves one so much in the dark. There is a small site plan, but none of the excavated areas is shown in any detail. The duration of the dig is unspecified, and there is no information about what proportion of the finds was made by the students and what by the builders. Several contributions are admirable, but others are thin on research, hypothesis and referencing. For example, the only photograph of the dig reveals a basement with a flat compacted floor, on which a tall person stands at full height with ease, despite the floor level having been raised by a metre of bones, soil and quicklime. But the section about the building's history, while addressing the cultural meaning of the Greek revival exterior at length, entirely overlooks the possibility that the original deep cellar may have been intended as a disposal area.

But the problem with the book is more fundamental than this. From title page to index it asserts and emphasizes the theme of racism in nineteenth-century medical training in the southern states. Before reading it I had no difficulty in accepting David Humphrey's 1973 finding that dissecting rooms in nineteenth-century America probably used a preponderance of black corpses. However, the archaeological data do not bear out the interpretation offered. Blakely and Judith M. Harrington extrapolate from their sample of 24 human tibias (which they believe to be representative) to say that 79% of those buried in the basement were of black origin. But Paul C. Dillingham, who explores the nutritional evidence using the same tibias (plus three more), finds almost equal numbers of European- and African-Americans. Moreover, he finds "no appreciable differences in the diet of African-Americans and Euro-Americans in the cadaver population".

The anthropologist Margaret Lock has observed that the objective of anthropology should be "not only to represent how others differ from us but also to use this knowledge to interrogate ourselves... what we should be doing is searching for the basis of our shared humanity". That the southern states were racist and that only better-off white men were educated as doctors in the nineteenth century was already known. If alongside the poor blacks buried in the basement there were also poor whites, and alongside males there were also females, why have we been offered a book that emphasizes only the racism of nineteenth-century medical training? □

Ruth Richardson is in the Department of Anatomy, University College London, Gower Street, London WC1E 6BT, UK.

G-protein diseases furnish a model for the turn-on switch

Taroh Iiri^{*†}, Zvi Farfel[‡] & Henry R. Bourne^{*}

^{*} Departments of Cellular and Molecular Pharmacology and Medicine, University of California, San Francisco, California 94143-0450, USA

[‡] Laboratory of Biochemical Pharmacology and Department of Medicine E, Sheba Medical Center, Tel Aviv University, Tel Hashomer, Israel

How does a trimeric G protein on the inside of a cell membrane respond to activation by a transmembrane receptor? G-protein mutations in patients with hypertension and inherited endocrine disorders enhance or block signals from stimulated receptors. In combination with three-dimensional crystal structures and results from biochemical experiments, the phenotypes produced by these mutations suggest a model for the molecular activation mechanism that relays hormonal and sensory signals transmitted by many transmembrane receptors.

Trimeric ($\alpha\beta\gamma$) G proteins relay signals from transmembrane receptors to intracellular enzymes and ion channels, thereby mediating vision, smell, taste and the actions of many hormones and neurotransmitters^{1,2}. Much effort has been devoted to elucidating the receptor-triggered 'turn-on' step of the GTPase cycle (Fig. 1): what happens at the atomic level when a receptor turns a G protein on, promoting exchange of GTP for GDP bound to the α subunit, followed by dissociation of α -GTP from $\beta\gamma$ and of both subunits from the receptor?

Figure 2 depicts a G-protein trimer in its probable orientation^{3,4} relative to a G-protein-coupled receptor and the plasma membrane. The receptor switch is thought to be composed of seven α -helices folded into a bundle that spans the membrane⁴ (reviewed in refs 5–8). We shall focus here on the crucial problem posed by the orientation of the $\alpha\beta\gamma$ trimer relative to the receptor (Fig. 2): cytoplasmic loops of most receptors (not depicted in the figure) are too short to reach more than halfway to the site where GDP is bound, about 30 Å from the plasma membrane^{1,8}; how then does the receptor act at a distance to cause release of bound GDP? We propose a speculative but testable model, in which the activating message is relayed from the receptor to the GDP-binding pocket of α by two complementary routes. Receptor-catalysed GDP/GTP exchange uses a switching mechanism and unique structural features that differ greatly from those that regulate GDP/GTP exchange in monomeric GTPases, such as Ras and the elongation factor (EF) Tu of protein synthesis. Just as G-protein mutations in human disease opened the way^{2,9–12} to understanding the G-protein 'turn-off' mechanism at the atomic level^{2,11–16}, other genetic diseases—ranging from rare endocrine disorders to hypertension—furnish critical clues to understanding the turn-on mechanism.

Action at a distance

Crystals have revealed three-dimensional (3D) structures of the substrate and product of the GDP/GTP exchange reaction (Fig. 1) catalysed by receptors. These are, respectively, the GDP-bound $\alpha\beta\gamma$ complex and the GTP-bound α and uncomplexed $\beta\gamma$ subunits^{3,15,17–22}. To understand the catalytic mechanism, however, we need a model of a G protein with an empty guanine-nucleotide-binding site ($\alpha\beta\gamma$ in Fig. 1). Such a model remains elusive because the key intermediate conformation is thermally labile in the absence of the catalyst, as might be expected. The catalyst—an activated receptor—does stabilize $\alpha\beta\gamma$, but receptors have proved hard to crystallize.

The 3D structure of α -GDP- $\beta\gamma$ (Fig. 2), the starting point of the GDP/GTP exchange reaction, reveals the 'action at a distance'

problem faced by the receptor. The guanine nucleotide (yellow) is cradled between two domains of the α subunit: one domain (grey) resembles those of Ras and other monomeric GTP-binding proteins; the other (orange) is an α -helical domain not found in other GTPases². Nucleotide-binding loops of the former domain, connecting β -strands and α -helices, share conserved amino-acid sequences with those of Ras and EF-Tu; loops at the opposite ends of the same β -strands and α -helices contact the receptor^{23–25}. $\beta\gamma$ binds to the Ras-like domain of α .

The complex (not shown) of two bacterial elongation (EF) factors, Tu and Ts, has provided the only available 3D structures^{26,27} of the 'empty nucleotide pocket' stage in a GDP/GTP exchange reaction. Figure 3 highlights, in blue and black, residues of α -GDP that correspond to those of EF-Tu that contact the exchange catalyst EF-Ts. The virtual footprint of Ts on α does not overlap with the α surface (red) available to the receptor^{23–25} (red). Instead, EF-Ts catalyses exchange by mounting a comprehensive attack on the nucleotide-binding pocket itself, poking a phenylalanine residue directly into it, releasing bound Mg^{2+} , and disrupting interactions of two loops with the β -phosphate and the guanine ring of GDP^{26,27}. Finally, Ts creates an exit route for the nucleotide by nudging yet a third loop out the way; this loop, cognate to the $\beta 3/\alpha 2$ loop of α , contains the black residues in Fig. 3.

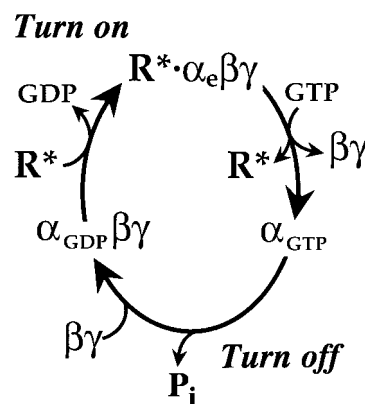


Figure 1 The GTPase cycle of trimeric G proteins. The 'turn-on' step begins when the activated receptor (R^*) associates with the trimer of (α -GDP- $\beta\gamma$), causing dissociation of GDP. Then GTP binds to the complex of R^* with the trimer in its 'empty' state ($\alpha\beta\gamma$), and the resulting GTP-induced conformational change causes α -GTP to dissociate from R^* and from $\beta\gamma$. After the 'turn-off' step (hydrolysis of bound GTP to GDP and inorganic phosphate, P_i), α -GDP reassociates with $\beta\gamma$.

[†] Present address: Fourth Department of Internal Medicine, University of Tokyo School of Medicine, 3-28-6 Mejirodai, Bunkyo-ku, Tokyo 112, Japan.

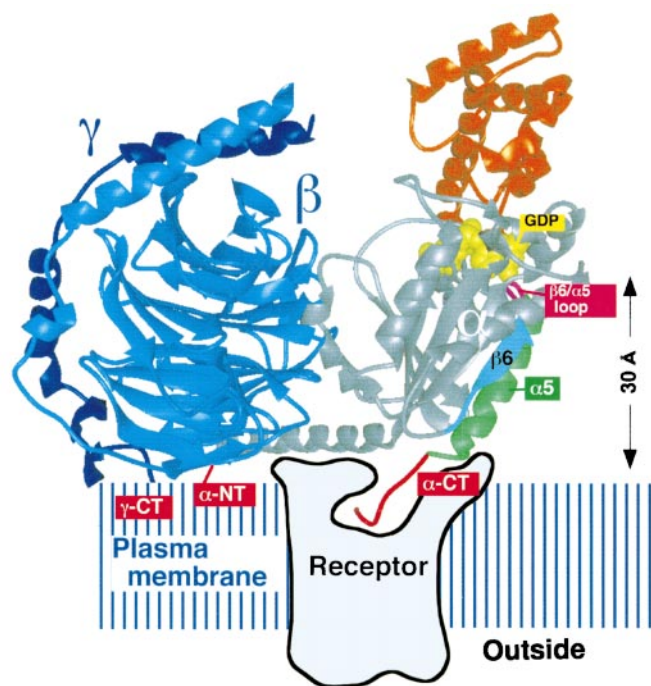


Figure 2 The postulated^{3,8} orientation of a G-protein trimer to a transmembrane receptor. The carboxy-terminal (CT) tail interacts with the receptor. Lipid modifications of the amino termini (NT) of $G\alpha$ and $G\gamma$ attach the trimer to the plasma membrane³. The $\beta 6$ strand and the $\alpha 5$ helix are postulated to transmit receptor-induced conformational change to the guanine ring of GDP, which contacts the $\beta 6/\alpha 5$ loop. The trimer structure is based upon that of G_i (ref. 3).

We propose (see below) that the receptor uses $G\beta\gamma$ to open a cognate exit route in $G\alpha$ during receptor-promoted GDP/GTP exchange. The footprint of $G\beta$ on $G\alpha$ (green and black in Fig. 3)^{3,21} partially overlaps the virtual footprint of EF-Ts; two cognate residues (black in Fig. 3) are touched by both proteins. A major part of the $G\beta\gamma$ -contacting surface of $G\alpha$ is composed of four consecutive secondary structures, including the 'switch 1' loop connecting the α -helical domain to the $\beta 2$ strand, the $\beta 2$ and $\beta 3$ strands, and the $\alpha 2$ helix. The 'lip' (pink in Fig. 3) of this contact surface, switch 1 and the $\beta 3/\alpha 2$ loop, occludes the exit route used by EF-Ts to release GDP from EF-Tu.

Regulation of GTP release by $G\alpha$ - and $G\beta\gamma$

The biochemical phenotype of one human $G\alpha$ mutant strikingly imitates the receptor-catalysed release of GDP. The patients, suffering from a combination of two rare endocrine diseases, carry a point mutation in codon 366 of the gene encoding the α -subunit of the stimulatory regulator of adenylyl cyclase, G_s (ref. 28). The alanine normally found at this position, located in the $\beta 6/\alpha 5$ loop (magenta in Fig. 2), makes a van der Waals contact with the guanine ring of the guanine nucleotide²². The slightly larger side chain of the serine substituted at this position (residue 366) causes the mutant α_s -A366S protein to release GDP spontaneously, at a rate 80 times faster than the wild-type subunit (α_s -WT). In the testis, rapid release of GDP mimics stimulation by gonadotropin receptors, accelerating GTP binding and G_s activation; the result, in males, is precocious puberty caused by autonomous production of testosterone (testotoxicosis). In other tissues, however, the patients show the diminished G_s -dependent hormone responses that are characteristic of type I pseudohypoparathyroidism. The defective hormone

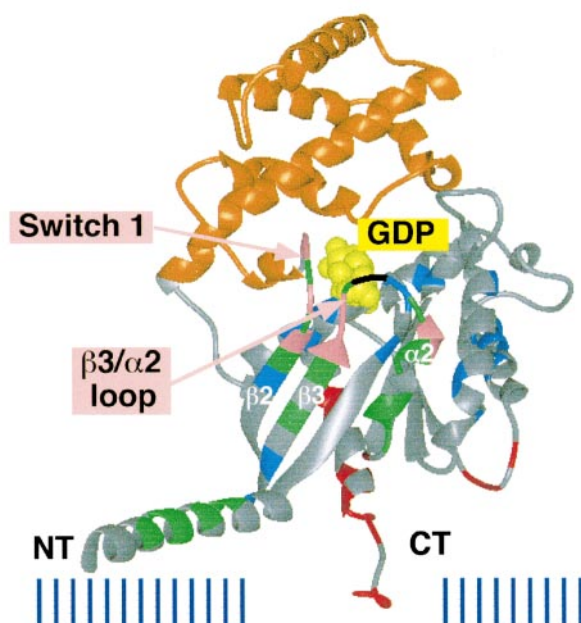


Figure 3 Surfaces of $G\alpha$ -GDP that may play a key role in the GDP/GTP exchange reaction. The $G\alpha$ -GDP component of the G_i trimer³ is rotated anticlockwise 90° about a vertical axis with respect to the orientation shown in Fig. 2; its $G\beta\gamma$ -binding surface faces the viewer. Coloured segments indicate amino acids that contact $G\beta\gamma$ (green)³, are thought to contact the receptor (red)²⁴, are cognate to the residues of EF-Tu contacted by its exchange catalyst EF-Ts (blue)²⁷, or serve as contact points for both $G\beta\gamma$ and EF-Ts (black). Pink regions demarcate the lip of a potential exit route for GDP, comprising the switch-1 loop (which connects the α -helical and Ras-like domains of $G\alpha$), the $\beta 3/\alpha 2$ loop, and part of the $\alpha 2$ helix.

responses result from the thermolability of α_s -A366S at body temperature (37 °C) and its stability at testis temperature (about 34 °C). The conserved alanine mutated in these patients is a potential control point for regulation of GDP release. Its replacement by other amino acids accelerates GDP release from other $G\alpha$ subunits ($G\alpha_{i2}$ and $G\alpha_x$; P. Wilson, J. Morales, T.I. and H.R.B., unpublished results) and also from Ras²⁹.

How could the receptor act at a distance to deform the $\beta 6/\alpha 5$ loop of $G\alpha$? Considerable evidence (from mutations, peptides and covalent modification by a bacterial toxin^{23,25,30,31}) points to the carboxy-terminal tail (red in Fig. 2) of $G\alpha$ as an important site for interaction with the receptor, achieved perhaps⁸ by insertion into a cavity³² formed by the seven-helix bundle (Fig. 2). The tail is located at the end of the $\alpha 5$ helix (green in Fig. 2)—that is, at the other end of the helix from the $\beta 6/\alpha 5$ loop. Receptors interact also with side chains of residues at the 'membrane end' of $\beta 6$ (cyan in Fig. 2)^{24,25,33,34}; like $\alpha 5$, $\beta 6$ connects the receptor-binding surface to the $\beta 6/\alpha 5$ loop. Moreover, release of GDP is accelerated by a truncation that removes part of the $\alpha 5$ helix and the C-terminal tail from $G\alpha_o$ (ref. 35). Thus $\beta 6$ and $\alpha 5$ probably bear at least partial responsibility for communicating the receptor signal to the nucleotide-binding pocket.

By analogy with EF-Ts acting on Tu, we and others¹² propose that $G\beta\gamma$ plays a second, complementary role in communicating the activating signal from receptor to the nucleotide-binding site. The scenario is simple: the membrane-facing side of the $G\alpha\beta\gamma$ complex contains a prominent cavity, previously noted^{3,8}, between $G\alpha$ and $G\beta\gamma$. The cavity provides an opportunity for loops of the activated receptor to tilt $G\beta\gamma$ and $G\alpha$ away from one another, causing $G\beta$ - $G\alpha$ contacts to pull the flexible lip away from a potential exit from

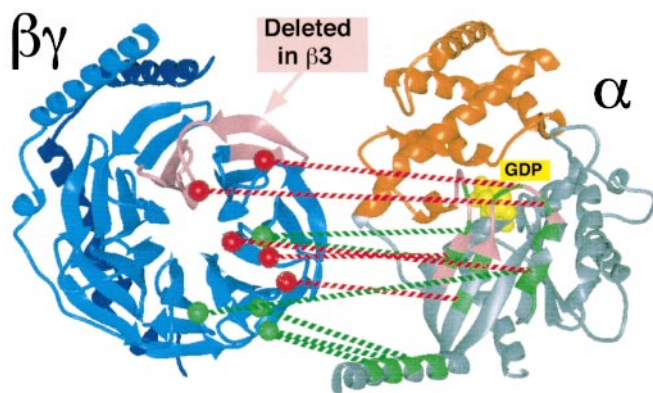


Figure 4 Contacts between $G\beta\gamma$ (left) and $G\alpha$ -GDP (right). With respect to their orientations in Fig. 2, $G\alpha$ -GDP and $G\beta\gamma$ are rotated about a vertical axis, anti-clockwise by 90° and clockwise by 60° , respectively. The C-terminal tail of $G\alpha$ -GDP and the contacts with the receptor and Ts are not shown; pink in $G\alpha$ indicates a 'lip' that occludes the postulated exit route for GDP (see text). In $G\beta$, pink indicates the stretch of amino acids deleted in $G\beta 3$ -s (ref. 37). The green and red balls indicate $G\beta$ positions where alanine substitutions reduce responsiveness of G_i to activation; of these, five (red balls) do not appreciably reduce the affinity of $G\alpha$ for binding $G\beta\gamma$ (see text)³⁶. Dashed lines connect residues at these mutated positions in $G\beta$ to the α -carbon(s) of the $G\alpha$ residue(s) with which each makes contact; red dashed lines indicate contacts that appear to be required for receptor activation but not for $G\alpha$ - $G\beta\gamma$ association; green dashed lines indicate contacts that are important for both functions³⁶.

the GDP-binding site. Thus the receptor uses $G\beta\gamma$ as a lever to effect GDP release at a distance.

Considerable evidence supports this scenario. Because receptors interact directly with $G\beta\gamma$ and require $G\beta\gamma$ for efficient activation of $G\alpha$ (ref. 8), it is not surprising that alanine substitutions at the $G\alpha$ - $G\beta$ interface^{24,36} impair receptor-induced GDP/GTP exchange. More instructive, five alanine substitutions in $G\beta$ inhibit receptor-promoted exchange but do not substantially interfere with binding of $G\beta\gamma$ to $G\alpha$ (ref. 36). This subset of $G\beta$ - $G\alpha$ contacts (red dashed lines in Fig. 4) probably plays important roles in receptor-induced GDP/GTP exchange, as suggested by their locations at or near the lip (pink in Figs 3 and 4) of the proposed exit route for GDP. In the unstimulated trimer, these contacts presumably stiffen the lip, enhancing affinity for GDP; a receptor-induced tilt of $G\beta\gamma$ away from $G\alpha$, however, could use the same contacts to pull the flexible lip away from the exit.

A recently reported human $G\beta$ mutation³⁷ produces a 'gain-of-function' G-protein signalling abnormality, perhaps by enhancing the proposed active role of $G\beta\gamma$ in receptor-catalysed activation of G proteins. Aberrant splicing of the transcript of a mutant gene that encodes the $\beta 3$ member of the $G\beta$ family³⁷, frequently seen in patients with essential hypertension, leads to production of a short protein, termed $G\beta 3$ -s, that lacks an internal stretch of 41 amino acids. Astonishingly, $G\beta 3$ -s functions in the signalling machinery of platelets and cultured cells from hypertensive patients and in insect cells expressing the mutant polypeptide³⁷. Expression of $G\beta 3$ -s is accompanied by enhanced sensitivity of G_i proteins to receptor activation.

The modular structure of $G\beta$ (Fig. 4) probably explains the surprising ability of $G\beta 3$ -s to fold and to transmit signals³⁸. $G\beta$ isozymes are composed of seven very similar β -sheets surrounding a central hole, like blades of a propeller^{19,21}. The aberrant splice excises the equivalent of one propeller blade (pink in Fig. 4), presumably allowing the six remaining blades to adopt a similar fold around the central hole. Despite its diminished girth, the new fold could preserve contacts with $G\gamma$, which stabilizes the $G\beta$ fold³⁹ and orients the $G\beta\gamma$ dimer with respect to the plasma membrane³.

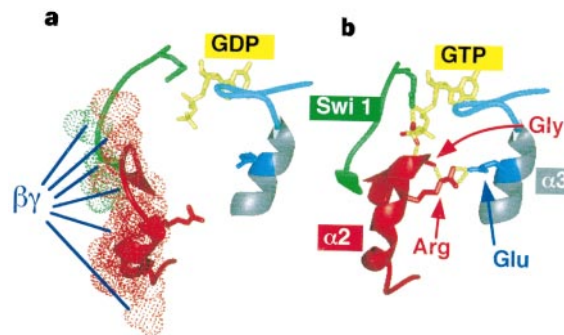


Figure 5 Different conformations of $G\alpha$ -GDP in the trimer³ and GTP-bound $G\alpha$ (ref. 18). **a**, In the trimer, association of $G\alpha$ -GDP with $G\beta\gamma$ induces movements of the switch 1 loop (Swi1 in **b**; green) and the $\alpha 2$ helix (red) away from the guanine nucleotide and places an arginine (red; see **b**) in the $\alpha 2$ helix out of reach of a glutamate (blue; see **b**) in the $\alpha 3$ helix; residues that contact $G\beta$ in the trimer are shown in a van der Waals representation (red and green dots). **b**, The GTP-bound $G\alpha$ conformation is stabilized by the intramolecular hasp (yellow dashes) formed by a salt bridge between the arginine residue (red) in the $\alpha 2$ helix and the glutamate (blue) in the $\alpha 3$ helix; the arginine side chain is also linked to the main-chain carbonyl of a conserved glycine in the $\beta 3/\alpha 2$ loop. The guanine nucleotide is shown in yellow, except that oxygens of the γ -phosphate of GTP are in red.

How might the $G\beta 3$ -s mutation enhance the transmission of conformational change from the receptor to the nucleotide-binding site? Excision of the propeller blade from $G\beta$ would change the positions of critical $G\alpha$ -contacting residues relative to one another and to the plane of the membrane. The $G\alpha$ contacts most affected would be in the lip that guards the proposed exit route for GDP (Fig. 4). This hypothesis can be tested biochemically.

A $G\alpha$ mutation disrupts GTP binding

Release of bound GDP is not enough to activate a G protein; in the second step of the GDP/GTP exchange reaction, GTP must enter the nucleotide-binding pocket of $\alpha\epsilon\beta\gamma$ and trigger the dissociation of $G\alpha$, $G\beta\gamma$ and the receptor (Fig. 1). This critical second step in the exchange reaction cannot be automatic, because the receptor must destabilize the guanine-nucleotide-binding site in order to promote release of bound GDP. Each of the complementary mechanisms we have proposed should destabilize binding of GTP as well as GDP; indeed, receptors can increase the rates of dissociation of GTP analogues from G proteins⁴⁰. Normally, however, GTP efficiently replaces GDP in the binding site because the γ -phosphate rescues $G\alpha$ from receptor-induced instability. The additional binding energy furnished by the γ -phosphate promotes a conformational change that causes $G\alpha$ to dissociate from $\beta\gamma$ and from the receptor; separation from the receptor definitively removes the destabilizing effect.

The essential moving part of this conformational switch is the $\alpha 2$ helix of $G\alpha$. In the trimer (Figs 2–4), many residues of this helix and the preceding $\beta 3/\alpha 2$ loop interact with $G\beta\gamma$ (refs 3, 21). In the transition from the trimer (Fig. 5a) to the $G\alpha$ -GTP conformation (Fig. 5b), the amino terminus of this helix moves about 3 Å closer to the guanine nucleotide; the helix also twists on its axis, exposing a different set of amino-acid side chains. Neither change could occur without dissociation of $G\alpha$ from $G\beta\gamma$, which stabilizes both the position and the axial rotation of the $\alpha 2$ helix in the trimer. Thus the transition from $R^*\alpha\epsilon\beta\gamma$ to $R^* + \alpha$ -GTP + $\beta\gamma$ (Fig. 1) reflects the outcome of a molecular tug-of-war, staged between $G\beta\gamma$ and the γ -phosphate of GTP, for controlling the position of the $\alpha 2$ helix. The

exact position of the $\alpha 2$ helix in $G\alpha$ -GTP (Fig. 5b) is specified by a link between an oxygen of the γ -phosphate and the main-chain amide of a conserved glycine in the $\beta 3/\alpha 2$ loop, which precedes the helix^{15,17,18}.

The link does not suffice for GTP to win the tug-of-war against $G\beta\gamma$, as shown by an instructive $G\alpha_s$ mutation^{41,42}, found in a family with type I pseudohypoparathyroidism. This mutation, substituting histidine for a conserved arginine at position 231 (ref. 41), breaks the elegant molecular device that normally gives the edge to GTP. Receptors appear to promote GDP release normally from α_s -R231H, but trigger binding of GTP at a rate 25-fold lower than to α_s -WT (ref. 42).

Comparing the 3D structure of the trimer to that of $G\alpha$ -GTP (Fig. 5) reveals the probable mechanism of the R231H activation defect. The arginine at position 231 (red in Fig. 5) is conserved in the $\alpha 2$ helix of all $G\alpha$ proteins. Upon binding of GTP, this helix moves towards the guanine nucleotide and twists about its axis to form a coordinated complex with residues in the $\alpha 3$ helix and the preceding loop^{17,18,22}. In this complex (Fig. 5b), the conserved arginine side chain forms a salt bridge with a conserved glutamate in the $\alpha 3$ helix, positioning $\alpha 2$ precisely with respect to $\alpha 3$; in addition, its guanidinium group stabilizes the main-chain oxygen of the same glycine, whose amide group interacts with the γ -phosphate of GTP. Mutations at either position of this conserved arginine–glutamate pair also inhibit activation of other trimeric G proteins^{24,43}. The α_s -R231H phenotype suggests that the salt bridge serves as an intramolecular hasp to fasten together the $\alpha 2$ and $\alpha 3$ helices⁴², allowing $G\alpha$ to hold GTP tightly and maintain the active conformation more effectively (Fig. 5b).

The hormone-response defect in patients who inherit the α_s -R231H mutation results from failure of $G\alpha_s$ to complete the second step of the GDP/GTP exchange reaction. The broken hasp weakens the ability of $G\alpha$ to stabilize the GTP-bound conformation required to disengage from $G\beta\gamma$ and the receptor. Consequently, GTP loses the tug-of-war against $G\beta\gamma$ for control of the $\alpha 2$ helix.

Perspective

We have outlined a speculative but comprehensive working model, in which the membrane-bound receptor uses two complementary mechanisms to act at a distance on the G protein's guanine-nucleotide-binding pocket. The model highlights unique features of the molecular mechanism crafted by evolution to regulate GDP/GTP exchange on trimeric G proteins. These proteins differ from their cousins, the monomeric GTPases, in many ways. Unlike $G\alpha$, the latter proteins show little or no preference for binding GTP over GDP, in part because they lack a hasp linking the $\alpha 2$ and $\alpha 3$ helices (as noted previously²⁷).

In creating $G\beta\gamma$ as an adjuvant catalyst of GDP/GTP exchange, evolution generated four other important consequences for signal transduction: (1) in the absence of hormone, $G\beta\gamma$ reduces signal noise by stabilizing GDP binding^{39,44,45}, even though $G\beta\gamma$ is also required for transducing the hormonal stimulus; (2) GTP-dependent release of free $G\beta\gamma$ provides a second potential regulator of downstream effectors, in addition to $G\alpha$ -GTP^{39,45}; (3) receptor-dependent activation of $G\alpha$ is irreversible, because the low affinity of $G\alpha$ -GTP for $G\beta\gamma$ prevents it from interacting with the receptor; (4) the tighter association of GTP than GDP with $G\alpha$ means that the transmitted signal cannot be terminated by dissociation of bound nucleotide, but only by its hydrolysis.

This model will certainly not prove correct in all its details. Eventually, G-protein-coupled receptors will be crystallized in their active and inactive forms and in association with $\alpha\beta\gamma$. In the interim, testing the model will require painstaking biochemical experiments with pure receptors and G-protein subunits. □

- Hamm, H. E. The many faces of G protein signaling. *J. Biol. Chem.* **273**, 669–672 (1998).
- Sprang, S. R. G protein mechanisms: Insights from structural analysis. *Annu. Rev. Biochem.* **66**, 639–678 (1997).

- Lambright, D. G. *et al.* The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature* **379**, 311–319 (1996).
- Baldwin, J. M., Schertler, G. F. & Unger, V. M. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* **272**, 144–164 (1997).
- Hargrave, P. A., Hamm, H. E. & Hofmann, K. P. Interaction of rhodopsin with the G-protein, transducin. *Bioessays* **15**, 43–50 (1993).
- Fahmy, K., Siebert, F. & Sakmar, T. P. Photoactivated state of rhodopsin and how it can form. *Biophys. Chem.* **56**, 171–181 (1995).
- Strader, C. D., Fong, T. M., Graziano, M. P. & Tota, M. R. The family of G-protein-coupled receptors. *FASEB J.* **9**, 745–754 (1995).
- Bourne, H. R. How receptors talk to trimeric G proteins. *Curr. Opin. Cell Biol.* **13**, 134–142 (1997).
- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: A conserved switch for diverse cell functions. *Nature* **348**, 125–132 (1990).
- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: Conserved structure and molecular mechanism. *Nature* **349**, 117–127 (1991).
- Bourne, H. R. The arginine finger strikes again. *Nature* **389**, 673–674 (1997).
- Bohm, A., Gaudet, R. & Sigler, P. B. Structural aspects of heterotrimeric G-protein signaling. *Curr. Opin. Biotechnol.* **8**, 480–487 (1997).
- Scheffzek, K. *et al.* The Ras–RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic *ras* mutants. *Science* **277**, 333–338 (1997).
- Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. GTPase mechanism of G proteins from the 1.7-Å crystal structure of transducin α -GDP- AlF_4^- . *Nature* **372**, 276–279 (1994).
- Coleman, D. E. *et al.* Structure of active conformations of $G_{i\alpha 1}$ and the mechanism of GTP hydrolysis. *Science* **265**, 1405–1412 (1994).
- Tesmer, J. J., Berman, D. M., Gilman, A. G. & Sprang, S. R. Structure of RGS4 bound to AlF_4^- activated $G_{i\alpha 1}$: stabilization of the transition state for GTP hydrolysis. *Cell* **89**, 251–261 (1997).
- Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. The 1.8 Å crystal structure of transducin α -GDP: structural determinants for activation of a heterotrimeric G-protein α -subunit. *Nature* **369**, 621–628 (1994).
- Noel, J. P., Hamm, H. E. & Sigler, P. B. The 2.2 Å crystal structure of transducin α -GTP γ S. *Nature* **366**, 654–663 (1993).
- Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. Crystal structure of a G_A protein $\beta\gamma$ dimer at 2.1 Å resolution. *Nature* **379**, 369–374 (1996).
- Mixon, M. B. *et al.* Tertiary and quaternary structural changes in $G_{i\alpha 1}$ induced by GTP hydrolysis. *Science* **270**, 954–960 (1995).
- Wall, M. A. *et al.* The structure of the G protein heterotrimer $G_{i\alpha 1}\beta_1\gamma_1$. *Cell* **83**, 1047–1058 (1995).
- Sunahara, R. K., Tesmer, J. J. G., Gilman, A. G. & Sprang, S. R. Crystal structure of the adenylyl cyclase activator, $G_{s\alpha}$. *Science* **278**, 1943–1947 (1997).
- Garcia, P. D., Onrust, R., Bell, S. M., Sakmar, T. P. & Bourne, H. R. Transducin- α C-terminal mutations prevent activation by rhodopsin: A new assay using recombinant proteins expressed in cultured cells. *EMBO J.* **14**, 4460–4469 (1995).
- Onrust, R. *et al.* Receptor and $\beta\gamma$ binding sites in the α subunit of the retinal G protein transducin. *Science* **275**, 381–384 (1997).
- Hamm, H. E. *et al.* Site of G protein binding to rhodopsin mapped with synthetic peptides from the α subunit. *Science* **241**, 832–835 (1988).
- Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S. & Leberman, R. The structure of the *Escherichia coli* EF-Tu- EF-Ts complex at 2.5 Å resolution. *Nature* **379**, 511–518 (1996).
- Wang, Y., Jiang, Y., Meyering-Voss, M., Sprinzl, M. & Sigler, P. B. Crystal structure of the EF-Tu- EF-Ts complex from *Thermus thermophilus*. *Nature Struct. Biol.* **4**, 950–956 (1997).
- Iiri, T., Herzmark, P., Nakamoto, J. M., Van Dop, C. & Bourne, H. R. Rapid GDP release from $G_{i\alpha}$ in patients with gain and loss of endocrine function. *Nature* **371**, 164–168 (1994).
- Feig, L. A. & Cooper, G. M. Relationship among guanine nucleotide exchange, GTP hydrolysis, and transforming potential of mutated *ras* proteins. *Mol. Cell. Biol.* **8**, 2472–2478 (1988).
- Schwindinger, W. F., Miric, A., Zimmerman, D. & Levine, M. A. A novel G_{α} mutant in a patient with Albright hereditary osteodystrophy uncouples cell surface receptors from adenylyl cyclase. *J. Biol. Chem.* **269**, 25387–25391 (1994).
- West, R. E., Moss, J., Vaughan, M., Liu, T. & Liu, T. Y. Pertussis toxin-catalyzed ADP-ribosylation of transducin. Cysteine-347 is the ADP-ribose acceptor site. *J. Biol. Chem.* **260**, 14428–14430 (1985).
- Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L. & Khorana, G. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **274**, 768–770 (1996).
- Bae, H. *et al.* Molecular determinants of selectivity in 5-hydroxytryptamine_{1B} receptor–G protein interactions. *J. Biol. Chem.* **272**, 32071–32077 (1997).
- Mazzoni, M. R. & Hamm, H. E. Interaction of transducin with light-activated rhodopsin protects it from proteolytic digestion by trypsin. *J. Biol. Chem.* **271**, 30034–30040 (1996).
- Denker, B. M., Schmidt, C. J. & Neer, E. J. Promotion of the GTP-liganded state of the $G_{\alpha o}$ protein by deletion of the C terminus. *J. Biol. Chem.* **267**, 9998–10002 (1992).
- Ford, C. E. *et al.* Molecular basis for interactions of G protein $\beta\gamma$ subunits with effectors. *Science* **280**, 1271–1274 (1998).
- Siffert, W. *et al.* Association of an allele of a human G-protein $\beta 3$ subunit gene with hypertension. *Nature Genet.* **18**, 45–48 (1998).
- Iiri, T. & Bourne, H. R. G proteins propel surprise. *Nature Genet.* **18**, 45–48 (1998).
- Clapham, D. E. & Neer, E. J. G protein $\gamma\beta$ subunits. *Annu. Rev. Pharmacol. Toxicol.* **37**, 167–203 (1997).
- Berstein, G. *et al.* Reconstitution of agonist-stimulated phosphatidylinositol 4,5-bisphosphate hydrolysis using purified m1 muscarinic receptor, G_{q11} , and phospholipase C- $\beta 1$. *J. Biol. Chem.* **267**, 8081–8088 (1992).
- Farfel, Z. *et al.* Pseudohypoparathyroidism, a novel mutation in the $\beta\gamma$ -contact region of $G_{i\alpha}$ impairs receptor stimulation. *J. Biol. Chem.* **271**, 19653–19655 (1996).
- Iiri, T., Farfel, Z. & Bourne, H. R. Conditional activation defect of a human $G_{i\alpha}$ mutant. *Proc. Natl Acad. Sci. USA* **94**, 5656–5661 (1997).
- Stratton, H. F., Zhou, J., Reed, S. I. & Stone, D. E. The mating-specific G_{α} protein of *Saccharomyces cerevisiae* downregulates the mating signal by a mechanism that is dependent on pheromone and independent of $G_{\beta\gamma}$ sequestration. *Mol. Cell. Biol.* **16**, 6325–6337 (1996).
- Higashijima, T., Ferguson, K. M., Sternweis, P. C., Smigel, M. D. & Gilman, A. G. Effects of Mg^{2+} on the $\beta\gamma$ -subunit on the interactions of guanine nucleotides with G proteins. *J. Biol. Chem.* **262**, 762–766 (1987).
- Sternweis, P. C. The active role of beta gamma in signal transduction. *Curr. Opin. Cell Biol.* **6**, 198–203 (1994).

Acknowledgements. This work was supported in part by a Julius Comroe Fellowship (T.I.), grants from the United States–Israel Binational Foundation and the Israeli Ministry of Health (Z.F.), and by NIH grants (H.R.B.).

Correspondence should be addressed to H.R.B. (e-mail: h_bourne@quickmail.ucsf.edu).

Magnetically mediated superconductivity in heavy fermion compounds

N. D. Mathur*, F. M. Grosche*, S. R. Julian, I. R. Walker, D. M. Freye, R. K. W. Haselwimmer & G. G. Lonzarich

Cavendish Laboratory and the Interdisciplinary Research Centre for Superconductivity, University of Cambridge, Cambridge CB3 0HE, UK

In a conventional superconductor, the binding of electrons into the paired states that collectively carry the supercurrent is mediated by phonons—vibrations of the crystal lattice. Here we argue that, in the case of the heavy fermion superconductors CePd_2Si_2 and CeIn_3 , the charge carriers are bound together in pairs by magnetic spin–spin interactions. The existence of magnetically mediated superconductivity in these compounds could help shed light on the question of whether magnetic interactions are relevant for describing the superconducting and normal-state properties of other strongly correlated electron systems, perhaps including the high-temperature copper oxide superconductors.

The ancient observation of action at a distance between magnetic materials such as lodestone inspired William Gilbert to postulate an analogous force between the Earth and the Sun¹. Later work showed that a parallel force holds atoms together. This charge–charge interaction is also ultimately responsible, in a subtle way, for the binding of Cooper pairs in superconductors. In traditional superconductors², this binding is most simply described in terms of the emission and absorption of waves of lattice density. Here we consider instead whether this binding could in some materials be most simply described in terms of the emission and absorption of waves of the electron spin density: that is, whether a type of bound state that was only identified this century can arise because of an effective spin–spin interaction (see, for example, refs 3–7) that is reminiscent of one of the oldest known forces.

The low-temperature properties of interacting electrons in conductors are normally described in terms of excited states known as ‘quasiparticles’^{8,9}. These entities are liable to be very different from bare electrons. Although they may obey the same ‘fermion’ quantum statistics, they can behave as though they are very ‘heavy’. We are interested in materials that support such quasiparticles, namely the ‘heavy fermion’ compounds. Exotic interactions can take place between quasiparticle charge carriers via their host medium, leading to new and subtle states of matter. One may imagine that, at each instant in its motion, a quasiparticle emits a wave that perturbs its environment in the manner of a stone entering a pond. Another charge carrier elsewhere can receive the signals, and thus charge carriers are able to communicate with each other. It is possible to tune the nature of these communications by altering the density of the crystal. In this way it is possible to alter dramatically the transport behaviour and other measured properties of a material. And in sufficiently calm conditions at low temperatures, bound states may form leading, in particular, to superconductivity.

We consider whether it is possible to tune the nature of charge carrier interactions such that magnetic coupling demonstrably dominates the more conventional non-magnetic channels of communication. In particular, we examine whether in such a scenario it would be possible for a magnetically mediated form of superconductivity to exist at sufficiently low temperatures. We present new evidence to suggest that this kind of superconductivity may indeed exist in some heavy fermion compounds, but that it is normally only viable extremely close to the critical lattice density n_c

at which long-range magnetic order is suppressed, and in specimens of extremely high purity. We discuss a magnetic-interactions model that provides a natural description of the temperature–density phase diagrams we have observed in two particularly simple heavy fermion compounds, namely cubic CeIn_3 and tetragonal CePd_2Si_2 . In each of these two systems, we find that both the variation of the magnetic and superconducting transition temperatures with lattice density, and the anomalous forms of the normal state resistivities, are qualitatively consistent with the magnetic-interactions model, which uses independently estimated parameters. This agreement, together with other features described here, provides compelling evidence for the probable existence of magnetically mediated superconductivity. Moreover, the magnetic interactions model, which may account for superconductivity in the materials studied, suggests how the properties of a material might be modified in order to increase the superconducting transition temperature. Interestingly, the changes required would lead us to materials with some of the features of the copper oxide superconductors.

The edge of magnetic order

We consider how to tune the charge carrier interactions so that the magnetic channel may become dominant. During the course of its motion, a quasiparticle nucleates various waves in the medium which affect other quasiparticles. The amplitudes of such waves depend on the strength of the coupling of a quasiparticle with the medium, and on the ease with which such waves can be excited in the medium. Magnetic waves of interest may be readily excited when a magnetic medium is near a critical lattice density, n_c , that renders it close to entering a state of continuous long-range magnetic order. Such waves in the non-magnetically ordered state are usually damped and may be described in terms of a relaxation frequency spectrum, Γ_q . And this quantity, which describes the characteristic rate of decay of a spontaneous fluctuation of the magnetization of wavevector $\mathbf{q}$, characterizes the dynamics of quasiparticle interactions (for a review, see ref. 10; see also refs 11–14).

In the simplest model, the retarded interaction produced by one quasiparticle on another depends on the product of three factors⁴, namely the relative orientation and magnitudes of the magnetic moments involved, $-\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2$, the square of a coupling parameter, and the space and time-dependent magnetic susceptibility, which itself depends on Γ_q . When it is dominant, this interaction can lead to two striking consequences. First, near n_c , magnetic waves tend to propagate over a long range. If the coupling parameter for such waves remains finite, the magnetic interactions are therefore also

* Present address: Department of Materials Science, University of Cambridge, Cambridge CB2 3QZ, UK (N.D.M.); MPI Chemische Physik fester Stoffe, Bayreuther Str. 40, 01187 Dresden, Germany (F.M.G.).

long-range. This can lead, as we shall see, to properties that are very different from those normally associated with metals, or with the so-called Fermi liquid model of the normal metallic state^{8,9}. Second, for some relative orientation of two quasiparticle moments, the interaction can be attractive in certain regimes of space and time. Hence bound states, and therefore superconductivity, may arise at low temperatures. In contrast to the classical superconductors, the bound states produced by magnetic interactions have orbital and perhaps spin angular momentum and are hence anisotropic in space. We also note that magnetic interactions tend to interfere with pairing via phonons and therefore are usually expected to suppress conventional forms of superconductivity in nearly magnetic metals^{15,16}. Further, we note that the coupling parameter may be non-local and that the response of the medium may be nonlinear. Hence in general magnetic interactions may not be described by the space-time susceptibility alone. The magnetic interactions model leads to a qualitative phase diagram of the form shown in Fig. 1.

The choice of materials

Evidence for the above phase diagram may be sought in compounds in which (1) the critical density n_c may be reached with readily accessible pressures (in our case <30 kbar); (2) the states are metallic on both sides of n_c with a single magnetic transition near n_c , or otherwise other interaction channels are liable to be important; (3) the magnetic transition is continuous at n_c , or otherwise the magnetic interactions would not be long-range nor necessarily dominant; and (4) pair-breaking effects, in particular impurity scattering, can be made sufficiently small.

Simple candidates include d metal ferromagnets^{17–24}, but these suffer from problems of purity, remoteness from n_c or, in some cases, pair-breaking due to lack of inversion symmetry. Similar complications arise in simple d metal antiferromagnets in which, moreover, the magnetic transitions are often not continuous. More promising candidates can be found among the f metals. We consider in particular the cerium-based heavy fermion systems.

To narrow the list down further, we examine the prediction of the magnetic interactions model for the superconducting transition temperature T_c . The solutions of the Eliashberg equations suggest that $T_c \sim \Gamma \theta \phi \exp(- (1 + \lambda)/g\lambda)$; (see ref. 6 for example). Here Γ is the bandwidth of the spectrum Γ_q , λ represents the enhancement of the quasiparticle mass due to magnetic interactions^{10–14} which

depends on a $\mathbf{q}$ -space average of $1/\Gamma_q$, and g measures the relative strength of the interaction in the given pair state. Nearly ferromagnetic metals are expected to produce pairs in spin-triplet states with an odd orbital angular momentum quantum number l , whereas nearly antiferromagnetic metals should produce spin-singlet states with l even (and normally greater than zero)^{3–7}. The factor θ depends on the form of Γ_q and, in part, represents damping due to incoherent inelastic scattering; ϕ represents damping due to scattering from impurities (either magnetic or non-magnetic for anisotropic pair states): it falls when there is an increase in the ratio of the superconducting coherence length ξ to the carrier mean free path l_{mfp} .

As it stands, this model for T_c predicts that superconductivity should be widespread in pure cerium-based heavy fermion compounds. But, despite an extensive search for two decades, superconductivity has been found at ambient pressure in these systems only in CeCu₂Si₂ (ref. 25). Here, T_c is insensitive to pressure and the pairing mechanism may be different from that described above^{26,27}. Indeed, the collapse of T_c upon reaching n_c in the closely related system CeCu₂Ge₂ under pressure²⁸ might well be described in terms of the pair breaking effects of magnetic interactions, as suggested in the early proposals regarding ambient pressure Pd (refs 15, 16).

The scarcity of superconductivity in the cerium heavy fermion metals may be due to impurities and competing channels. Sufficiently close to n_c however, magnetic interactions become long-range and can dwarf competing channels, at least for pair states of sufficiently high l . Moreover, because the quasiparticle mass m^* grows as n_c is approached^{10–14}, ξ tends to a minimum and hence ϕ tends to a maximum—at least if l_{mfp} does not fall too rapidly as $n \rightarrow n_c$. On the other hand, θ drops near n_c , and may indeed collapse in the nearly ferromagnetic metals⁵. This effect is less extreme in nearly antiferromagnetic metals and T_c is a maximum at n_c (ref. 29). We therefore confine our search to antiferromagnetic heavy fermion systems.

We estimate that in contrast to the d metals, it is possible to produce sufficiently pure cerium-based systems because ξ tends to be short. After an extensive survey, it was decided that CePd₂Si₂ and CeIn₃ were among the materials most likely to satisfy all the above criteria. CePd₂Si₂ crystallizes in a body-centred tetragonal structure like CeCu₂Si₂ and is, at ambient pressures, an antiferromagnet with $\mathbf{Q}$ along [110] below a Néel temperature T_N of ~ 10 K (refs 30–33). CeIn₃ forms a simple cubic structure and at ambient pressure orders in a classical antiferromagnetic state with $\mathbf{Q}$ along [111] below $T_N \sim 10.1$ K (refs 34–36). In both systems, the critical density n_c (where T_N vanishes) can be reached with an applied pressure in the range 25 to 30 kbar (refs 33, 36).

Measurements of CePd₂Si₂ and CeIn₃

High-purity samples of CePd₂Si₂ were synthesized as follows. Reactants (Ce 99.999%, Pd 99.9985%, Si 99.999%) were heated by the non-invasive technique of radiofrequency induction on a water-cooled copper boat in an ultrahigh-vacuum system which could be pressurized to up to 5 atmospheres with ultrahigh-purity argon after *in situ* purification with a hot cerium getterer (the starting purity was better than 1 p.p.m.). Our a -axis resistivity measurements were made on a - b plane single-crystal platelets from the resulting annealed ingot, the best of which had a residual resistivity of a few $\mu\Omega\text{cm}$, which is a factor of five below values previously reported. Electron microprobe analysis revealed neither any evidence of inhomogeneities (to a resolution of 0.1%), nor the presence of a second phase. A similar procedure was employed in the synthesis of CeIn₃.

Single-crystal resistivity measurements under pressure were made by a low-power a.c. four-terminal method within conventional BeCu/maraging steel piston-cylinder cells, filled with an equal mixture of *iso*-pentane and *n*-pentane¹⁷. The pressure was determined to ± 0.1 kbar from the superconducting transition of Sn, and the temperature via Ge and RhFe sensors placed on a Cu enclosure

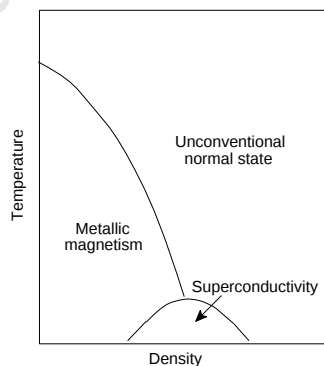


Figure 1 Possible temperature–density phase diagram of a pure metal in which magnetic order is quenched gradually with increasing lattice density. Near the critical density n_c , where the magnetic transition temperature vanishes, magnetic interactions become strong and long-range. The normal state is expected to be anomalous here and, at sufficiently low temperatures, it is expected to give way to a kind of superconductivity in which Cooper pairs are bound together by a glue of magnetic origin. Superconductivity may exist only over a very narrow range of densities near n_c —which is where magnetic interactions overwhelm other channels. Moreover, superconductivity may exist only in samples in which the carrier mean free path exceeds the superconducting coherence length. In most cases this requires samples of very high purity.

in good thermal contact with both the pressure cell and the sample leads. To try to ensure adequate pressure and temperature homogeneity, a slow cooling rate from room temperature, typically 0.2 K min^{-1} , was employed. Measurements were carried out from room temperature to the millikelvin temperature range within a pumped ^4He cryostat, an adiabatic demagnetization refrigerator and a top-loading dilution refrigerator.

Our key experimental results are summarized in Figs 2, 3. In the materials studied, the antiferromagnetic ordering temperature T_N , at which there is a discontinuity in the gradient of the resistivity ρ (not shown), was found to decrease slowly and monotonically with increasing pressure, p . Over a wide region of the CePd_2Si_2 phase diagram, T_N is close to being linear in p , and extrapolation from this regime to absolute zero allows us to define an effective critical pressure p_c of 28 kbar. In CeIn_3 , the variation of T_N with p is more rapid, and we estimate p_c to be ~ 26 kbar. The behaviour of T_N as it falls below 1 K has not been resolved in these studies.

In the case of CePd_2Si_2 , the resistivity ρ does not exhibit the standard T^2 form expected of a Fermi liquid. Careful analysis shows that near p_c it in fact varies as $T^{1.2 \pm 0.1}$ over nearly two decades in temperature down to the millikelvin range (Fig. 2, inset). Below 500 mK and in a narrow region near p_c , we observe an abrupt drop in ρ to below the detection limit, consistent with the occurrence of a superconducting transition, as discovered during our initial observations in September 1994^{37,38}. At a given pressure, this transition may be characterized by a temperature T_c , at which ρ falls to 50% of its normal state value. The width of this transition grows markedly as the pressure is varied away from p_c . We stress that experimentally, ρ is found to actually vanish only close to p_c . By energizing a Nb–Ti coil placed in our pressure cell, it was established that the upper critical field B_{c2} varies as $dB_{c2}(p_c)/dT \sim -6 \text{ T/K}$ near T_c . This is a high rate of change for such a small value of T_c —much higher than the expected figure for a conventional superconductor. However, it is the same order of magnitude as the value found in the heavy fermion superconductor CeCu_2Si_2 (ref. 27). We note that in a traditional analysis, the slope of $B_{c2}(T)$ at T_c implies a superconducting coherence length of $150 \text{ \AA}$, a value which is below the value of l_{mfp} that we estimate for our best samples. No superconductivity has been observed in specimens with residual resistivities above several $\mu\Omega \text{ cm}$, namely those with an estimated l_{mfp} that is

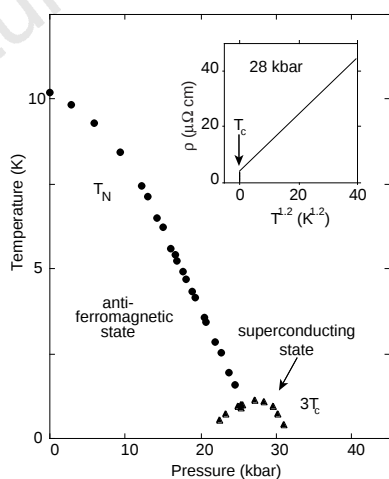


Figure 2 Temperature–pressure phase diagram of high-purity single-crystal CePd_2Si_2 . Superconductivity appears below T_c in a narrow window where the Néel temperature T_N tends to absolute zero. Inset: the normal state ρ -axis resistivity above the superconducting transition varies as $T^{1.2 \pm 0.1}$ over nearly two decades in temperature^{27,30}. The upper critical field B_{c2} at the maximum value of T_c varies near T_c at a rate of approximately -6 T/K . For clarity, the values of T_c have been scaled by a factor of three, and the origin of the inset has been set at 5 K below absolute zero.

substantially below ξ (for a similar example, see ref. 39).

In the case of cubic CeIn_3 , we find that very close to p_c the normal state resistivity assumes a non-Fermi liquid form, but this time^{40,41} varies as $T^{1.6 \pm 0.2}$. Thus, near their respective critical pressures, the resistivity exponent in the cubic material is significantly higher than it is in tetragonal CePd_2Si_2 . In a very narrow region near p_c , we again see a sharp drop in ρ to below the detection limit, but at somewhat lower temperatures than the transitions observed in CePd_2Si_2 . This is consistent with the occurrence of superconductivity in yet another cerium compound on the edge of long-range magnetic order^{40,41}.

We stress that in each material studied, both the form of the temperature dependence of the normal state resistivity, and the nature and existence of the superconducting transition are sensitive to sample quality. In particular, the superconducting transitions appear only in samples with residual resistivities in the low $\mu\Omega \text{ cm}$ range, as expected in the case of anisotropic pair states with coherence lengths of the order of a few hundred ångströms.

Magnetic interactions

The observed temperature–pressure phase diagrams for both CePd_2Si_2 and CeIn_3 are at least qualitatively consistent with what is expected in terms of the magnetic interactions model (Fig. 1). We now consider a more quantitative comparison. In the following it is assumed that the magnetic transition is continuous and that n is close to n_c . The incoherent scattering of quasiparticles via magnetic interactions is then expected to lead to a resistivity of the form

$$\rho = \rho_0 + AT^x \quad (1)$$

where ρ_0 and A are constants and the exponent x is smaller than two, that is, smaller than it is in a conventional Fermi liquid at low T

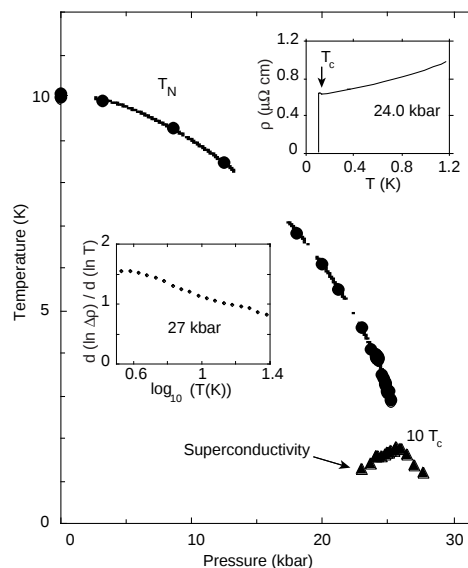


Figure 3 Temperature–pressure phase diagram of high-purity single-crystal CeIn_3 . A sharp drop in the resistivity consistent with the onset of superconductivity below T_c is observed in a narrow window near p_c , the pressure at which the Néel temperature T_N tends to absolute zero. Upper inset: this transition is complete even below p_c itself. Lower inset: just above p_c , where there is no Néel transition, a plot of the temperature dependence of $d(\ln \Delta\rho)/d(\ln T)$ is best able to demonstrate that the normal state resistivity varies as $T^{1.6 \pm 0.2}$ below several degrees K (ref. 29) ($\Delta\rho$ is the difference between the normal state resistivity and its residual value—which is calculated by extrapolating the normal-state resistivity to absolute zero). For clarity, the values of T_c have been scaled by a factor of ten. The resistivity exponents of CeIn_3 and CePd_2Si_2 may be understood by taking into account the underlying symmetries of the antiferromagnetic states and using the magnetic interactions model. Superconductivity near n_c in pure samples is expected to be a natural consequence of the same model.

(refs 10, 11, 14, 42). For a three-dimensional ferromagnet, $x = 5/3$, in agreement with several nearly ferromagnetic d metals^{17,21}. For an antiferromagnet, $x = 3/2$ in three dimensions and $x = 1$ in two dimensions^{11,14}, provided that in both cases ρ_0 and T are not too low⁴³. The former is indeed consistent with our findings in cubic CeIn₃, whereas the latter is closer to tetragonal CePd₂Si₂. The quasi two-dimensional character of the spin-relaxation spectrum of CePd₂Si₂ may be anticipated from the observed [110] direction of the ordering wavevector $\mathbf{Q}$ observed at ambient pressure. As shown in Fig. 4, the spin configuration for this case implies that the nearest-neighbour spin coupling along the c direction is frustrated. This leads to a weaker variation of $\Gamma_{\mathbf{Q}+\mathbf{q}}$ when $\mathbf{q}$ is along the c axis than when it lies in the basal plane, and hence to a quasi two-dimensional behaviour for the magnetic interactions. We note that along the c axis, if a slow q^4 variation in $\Gamma_{\mathbf{Q}+\mathbf{q}}$ survives in an even-power expansion of $\Gamma_{\mathbf{Q}+\mathbf{q}}$ in q , then the effective dimensionality is $5/2$ and $x = 5/4$. In this way, the exponent $3/2 > x > 1$ observed in CePd₂Si₂ can be interpreted naturally. But because of the reduced dimensionality, an elementary treatment may in this case be insufficient.

The magnetic interactions model also predicts that the magnetic transition temperature collapses as $|p - p_c|^\gamma$ near p_c (refs 10–14). For a three-dimensional ferromagnet, the exponent $\gamma = 3/4$, in agreement as before with observations in the d metals^{17,21}. For an antiferromagnet, $\gamma = 2/3$ in three dimensions, and $\gamma = 1$ in two dimensions. The quasi-linear variation of $T_N(p)$ in CePd₂Si₂ and the more rapid $T_N(p)$ in CeIn₃ near p_c are not inconsistent with these predictions.

At sufficiently low temperatures, the incoherent scattering that yields $\rho(T)$ is expected to go over to coherent scattering if the quasiparticles attract. This can lead to bound states and superconductivity. We now examine the predicted order of magnitude of T_c . Near n_c , magnetic interactions are expected to dominate other channels and for our samples, $\phi \approx 1$. In these circumstances, the numerical solutions of the Eliashberg equations for the T_c of antiferromagnets given in ref. 29 may be relevant. At n_c , the expected T_c depends strongly on Γ , which may be estimated from bulk properties and neutron-scattering data^{30–32,34–36}. In both CeIn₃ and CePd₂Si₂, we find Γ to be 30–50 K near p_c . Ref. 29 then suggests maximum values of T_c of a few tenths of a degree K at n_c , in order of magnitude agreement with our observations. With the same parameters, we can also understand the order of magnitude of A in equation (1) (ref. 41).

The estimates of T_c apply only very close to n_c . In impure samples, the extremely rapid fall of $T_c(p)$ may be due to a drop in ϕ , whereas in pure samples, an anticipated sharp fall in $g\lambda$ and the contribution of other interaction channels may take over. These effects may naturally account for the difficulty of finding magnetically mediated superconductors.

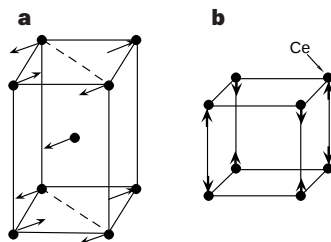


Figure 4 The positions and spin alignments of the Ce atoms in the unit cells of the two materials studied. **a**, The body-centred tetragonal unit cell of CePd₂Si₂ has the antiferromagnetic ordering wavevector $\mathbf{Q}$ along [110]. The Pd and Si atoms are not shown. Note that, by symmetry, the magnetic couplings along the c axis between nearest neighbours would actually vanish in this spin configuration. **b**, The simple cubic unit cell of CeIn₃ has $\mathbf{Q}$ along [111]. The In atoms (not shown) are located at the centre of the faces of the cubic unit cell.

Link to other highly correlated fermion systems?

Further evidence for the existence of magnetically mediated superconductivity in cerium heavy fermion compounds arose when similar findings were reported in CeRh₂Si₂ (ref. 44) and CeCu₂ (ref. 45). It is possible that these materials may also be understood in terms of the magnetic interactions model discussed here. But a detailed analysis is not yet possible. The metallurgical properties of CeRh₂Si₂ and CeCu₂ are complex and the transitions near n_c are more abrupt than those in CePd₂Si₂ or CeIn₃. Moreover, in CeRh₂Si₂ and possibly CeCu₂ as well, there exist two or more magnetic transitions that considerably complicate theoretical modelling. More recently, superconductivity has also been observed in CeNi₂Ge₂^{46,47}. This system has the same lattice and electronic structure as CePd₂Si₂ and behaves at ambient pressure much as CePd₂Si₂ does at p_c . Thus, there is increasing evidence, since the early observations in CePd₂Si₂ and CeRh₂Si₂, that superconductivity may occur quite generally in the immediate vicinity of n_c , in samples of sufficiently high quality, at least when the magnetic transition is continuous.

These considerations may also be important in the uranium-based heavy fermion superconductors^{48–53}. But in some of these systems, magnetic interactions may compete with others that are associated with incipient quadrupolar or higher multipolar ordering^{52,53}. Similarly, superconductivity observed on the border of metal–insulator Mott transitions may also originate from magnetic binding^{9,54–56}. But the magnetic transitions may not be continuous and moreover, subtle interactions associated with incipient localization could be important in these systems^{9,56}. Magnetic couplings are also thought to arise in liquid ³He and play a key role in the formation of the triplet p -wave superfluid state below a few millikelvin (refs 3, 4, 57). Even in this relatively simple fermionic system, however, the quasiparticle interactions may be complex and, in particular, may reflect the tendency of the quantum liquid to order in magnetic and spatially localized states.

It is interesting to consider how the low transition temperatures of the cerium based heavy fermion superconductors might be increased within the framework of the present magnetic pairing model. The formula presented earlier for T_c suggests that it is necessary to increase the bandwidth Γ of the magnetic relaxation spectrum without simultaneously forcing a drastic reduction in the parameter λ which appears in the exponent. The bandwidth may be increased by going over from narrow f band materials to broader d band systems—and λ , which depends on a momentum–space average of $1/\Gamma_{\mathbf{q}}$ (refs 10–14), may be prevented from dropping excessively by reducing the dispersion along one axis, by reducing the effective dimensionality. This also has the advantage of generally increasing g , particularly in the anisotropic singlet case. In addition, we note that for spin 1/2 particles, the magnitude of $\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2$, which appears in the formula for T_c , has a quantum average that is three times larger in the singlet state than it is in the triplet state. Both this fact, and the effects of incoherent scattering, tend to favour antiferromagnetic rather than ferromagnetic interactions.

The sum of these considerations would lead us to look for high superconducting transition temperatures in d metal compounds of reduced dimensionality on the border of antiferromagnetic order. It is interesting that these properties are in fact similar to those of the copper oxide superconductors. This may prove to be a coincidence, as other factors beyond those given here may be of comparable or even greater importance. Nevertheless, the search for a possible interpretation of the behaviour of the copper oxides in terms of some sort of magnetic-interactions model seems reasonable on the basis of the findings summarized here. Several scenarios with similar starting outlooks to the one we describe^{3–7,65–67} have been put forward^{58–61,68}, but even more exotic models may be required. As yet there is no general agreement on the best way of thinking about the copper oxides, but it is interesting that their spin liquid properties^{9,69}, and—in common with our phase diagrams—the

interplay of antiferromagnetism and superconductivity⁶², are persistent and increasingly more frequent themes.

'Magnetic glue'

The superconductivity we have observed in the metals CePd₂Si₂ and CeIn₃ is restricted to high-quality samples at the very edge of magnetic order. We believe that this is because it occurs only if there are attractive magnetic interactions that are strong enough to overcome competing interactions, and if it is possible for these magnetic interactions to create Cooper pairs that are small enough effectively not to see impurities. To support this statement, we argue that within the context of a single model based on magnetic interactions, it is possible to describe the key features of the phase diagrams and the anomalous normal state resistivities of the two materials in a consistent manner.

Therefore, we suggest that our evidence speaks strongly in favour of the existence of a superconducting state which arises because of magnetism rather than in spite of it. We imagine that the charge carriers in this superconducting state are held together in pairs by a 'magnetic glue'. Although magnetically mediated superconductivity is hard to prove because conditions must be just right in order to observe and infer its existence, we do not rule out the possibility that it is widespread. Thus our findings could shed some further light on current models of magnetic coupling in other strongly correlated electron systems—extending perhaps even to the high-temperature superconductors. If there is an alternative interpretation of our experimental results, it would probably involve even more unconventional thinking.

Note added in proof: Evidence in support of the 5/2 dimensional model discussed for CePd₂Si₂ (ref. 46) has been found via recent inelastic neutron scattering⁶³ in CeCu_{6-x}Au_x (ref. 64). In contrast to CePd₂Si₂ and CeIn₃, this system is disordered and hence does not exhibit the form of the resistivity or the superconductivity discussed in this Article. □

Received 29 January; accepted 29 April 1998.

1. Gilbert, W. *De Magnete, Magneticisque Corporibus et de Magno Magnets Tellure* (London 1600).
2. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
3. Brinkman, W. F., Serene, J. W. & Anderson, P. W. Spin-fluctuation stabilization of anisotropic superfluid states. *Phys. Rev. A* **10**, 2386–2394 (1974).
4. Leggett, A. J. A theoretical description of the new phases of liquid ³He. *Rev. Mod. Phys.* **47**, 331–414 (1975).
5. Fay, D. & Appel, J. Coexistence of *p*-state superconductivity and itinerant ferromagnetism. *Phys. Rev. B* **22**, 3173–3182 (1980).
6. Millis, A. J., Sachdev, S. & Varma, C. M. Inelastic-scattering and pair breaking in anisotropic and isotropic superconductors. *Phys. Rev. B* **37**, 4975–4986 (1988).
7. Bickers, N. E., Scalapino, D. J. & White, S. R. Conserving approximations for strongly correlated electron-systems—Bethe–Salpeter equation and dynamics for the two-dimensional Hubbard model. *Phys. Rev. Lett.* **62**, 961–964 (1989).
8. Landau, L. D. in *Collected Papers* Ch. 90, 91 (Gordon & Breach, New York, Pergamon, Oxford, 1965).
9. Anderson, P. W. *Theory of Superconductivity of High-T_c Cuprate Superconductors* (Princeton Univ. Press, 1997).
10. Lonzarich, G. G. *Electron* (ed. Springford, M.) Ch. 6 (Cambridge Univ. Press, 1997).
11. Moriya, T. *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985).
12. Hertz, J. A. Quantum critical phenomena. *Phys. Rev. B* **14**, 1165–1184 (1976).
13. Millis, A. J. Effect of a non-zero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
14. Moriya, T. & Takimoto, T. Anomalous properties around magnetic instability in heavy-electron systems. *J. Phys. Soc. Jpn.* **64**, 960–969 (1995).
15. Berk, N. F. & Schrieffer, J. R. Effect of ferromagnetic spin correlations on superconductivity. *Phys. Rev. Lett.* **17**, 433–435 (1966).
16. Gladstone, G., Jensen, M. A. & Schrieffer, J. R. in *Superconductivity* (ed. Parks, R. D.) Vol. 2 (Dekker, New York, 1969).
17. Pfeleiderer, C., McMullan, G. J., Julian, S. R. & Lonzarich, G. G. Magnetic quantum phase transition in MnSi under hydrostatic pressure. *Phys. Rev. B* **55**, 8330–8338 (1997).
18. Ishikawa, Y., Noda, Y., Uemura, Y. J., Majkrzak, C. F. & Shirane, G. Paramagnetic spin fluctuations in the weak itinerant-electron ferromagnet MnSi. *Phys. Rev. B* **31**, 5884–5893 (1985).
19. Lonzarich, G. G. & Taillefer, L. J. Effect of spin fluctuations on the magnetic equation of state of ferromagnetic or nearly ferromagnetic metals. *Phys. C: Solid State Phys.* **18**, 4339–4371 (1985).
20. Lonzarich, G. G. The magnetic equation of state and heat capacity in weak itinerant ferromagnets. *J. Magn. Magn. Mater.* **54–57**, 612–616 (1986).
21. Grosche, F. M., Pfeleiderer, C., McMullan, G. J., Lonzarich, G. G. & Bernhoeft, N. R. Critical behaviour of Zn₂. *Physica B* **206 & 207**, 20–22 (1995).
22. Bernhoeft, N. R., Lonzarich, G. G., Mitchell, P. W. & McK. Paul, D. Magnetic excitations in Ni₂Al at low energies and long wavelengths. *Phys. Rev. B* **28**, 422–424 (1983).
23. Bernhoeft, N. R., Law, S. A., Lonzarich, G. G. & McK. Paul, D. Magnetic excitations in Zn₂ at low energies and long wavelengths. *Physica Scripta* **38**, 191–193 (1988).

24. Lonzarich, G. G., Bernhoeft, N. R. & McK. Paul, D. Spin density fluctuations in magnetic metals. *Physica B* **156 & 157**, 699–705 (1989).
25. Steglich, F. *et al.* Superconductivity in the presence of strong Pauli paramagnetism: CeCu₂Si₂. *Phys. Rev. Lett.* **43**, 1892–1896 (1979).
26. Razafimandimby, H., Fulde, P. & Keller, J. On the theory of superconductivity in Kondo lattice systems. *Z. Phys. B-Cond. Matter* **54**, 111–120 (1984).
27. Cox, D. L. Selection-rules for 2-channel Kondo models of U⁴⁺ and Ce³⁺ ions in metals. *Physica B* **188**, 312–316 (1993).
28. Jaccard, D., Behnia, K. & Sierro, J. Pressure-induced heavy fermion superconductivity of CeCu₂Ge₂. *Phys. Lett.* **163A**, 475–480 (1992).
29. Nakamura, S., Moriya, T. & Ueda, K. Spin fluctuation-induced superconductivity in two- and three-dimensional systems. *J. Phys. Soc. Jpn.* **65**, 4026–4033 (1996).
30. Grier, B. H., Lawrence, J. M., Murgai, V. & Parks, R. D. Magnetic ordering in CeM₂Si₂ (M = Ag, Au, Pd, Rh) compounds as studied by neutron diffraction. *Phys. Rev. B* **29**, 2664–2672 (1984).
31. Grier, B. H., Lawrence, J. M., Horn, S. & Thompson, J. D. Inelastic magnetic neutron scattering in CeT₂Si₂ (T = Au, Pd, Ru or Pt): crystal fields and quasi-elastic scattering. *J. Phys. C: Solid State Phys.* **21**, 1099–1110 (1988).
32. Hippert, F., Hennion, B., Mignot, J.-M. & Lejay, P. Magnetic excitations in the antiferromagnetic Kondo compound CePd₂Si₂. *J. Magn. Magn. Mater.* **108**, 177–178 (1992).
33. Thompson, J. D., Parks, R. D. & Borges, H. Effect of pressure on the Néel temperature of Kondo-lattice systems. *J. Magn. Magn. Mater.* **54–57**, 377–378 (1986).
34. Lawrence, J. Speculations on the critical behavior of reduced moment anti-ferromagnetic cerium compounds. *J. Appl. Phys.* **53**, 2117–2121 (1982).
35. Lassailly, Y., Burke, S. K. & Flouquet, J. Strong Q-dependence of the magnetic susceptibility of CeIn₃. *J. Phys. C: Solid State Phys.* **18**, 5737–5747 (1985).
36. Flouquet, J. *et al.* Magnetic instability in Ce heavy fermion compounds. *J. Magn. Magn. Mater.* **90 & 91**, 377–382 (1990).
37. Grosche, F. M., Julian, S. R., Mathur, N. D. & Lonzarich, G. G. Magnetic and superconducting phases in CePd₂Si₂. *Physica B* **224**, 50–52 (1996).
38. Julian, S. R. *et al.* The normal states of magnetic *d* and magnetic *f* transition metals. *J. Phys. Cond. Matter* **8**, 9675–9688 (1996).
39. Mackenzie, A. P. *et al.* Extremely strong dependence of superconductivity on disorder in Sr₂RuO₄. *Phys. Rev. Lett.* **80**, 161–164 (1998).
40. Walker, I. R., Grosche, F. M., Freye, D. M. & Lonzarich, G. G. The normal and superconducting states of CeIn₃ and CePd₂Si₂ on the border of antiferromagnetic order. *Physica C* **282**, 303–306 (1997).
41. Grosche, F. M. *et al.* Superconductivity and anomalous normal states in cerium heavy fermion compounds. Preprint.
42. Althuler, B. L., Ioffe, L. B. & Millis, A. J. Low-energy properties of fermions with singular interactions. *Phys. Rev. B* **50**, 14048–14064 (1994).
43. Hlubina, R. & Rice, T. M. Resistivity as a function of temperature for models with hot-spots on the fermi-surface. *Phys. Rev. B* **51**, 9253–9260 (1995).
44. Movshovich, R. *et al.* Superconductivity in heavy-fermion CeRh₂Si₂. *Phys. Rev. B* **53**, 8241–8244 (1996).
45. Vargoz, E., Link, P. & Jaccard, D. CeCu₂: A new heavy-fermion superconductor near 6 GPa. *Physica B* **230**, 182–185 (1997).
46. Grosche, F. M. *et al.* Superconductivity and anomalous normal state in CePd₂Si₂ and CeNi₂Ge₂. Preprint aps 1997 aug 27_001 available at (<http://publish.aps.org/eprint/>).
47. Steglich, F. *et al.* Are heavy-fermion metals Fermi liquids? *Z. Phys. B* **103**, 235–242 (1997).
48. Ott, H. R., Rudigier, H., Fisk, Z. & Smith, J. L. UBe₁₃—an unconventional actinide superconductor. *Phys. Rev. Lett.* **50**, 1595–1598 (1983).
49. Stewart, G. R. Heavy-fermion systems. *Rev. Mod. Phys.* **56**, 755–787 (1984).
50. Fisk, Z. *et al.* Heavy-electron metals: new highly correlated states of matter. *Science* **239**, 33–42 (1988).
51. Grewe, N. & Steglich, F. in *Handbook of the Physics and Chemistry of the Rare Earths* (eds Gschneidner, K. A. Jr. & Eyring, L. L.) **14**, 343 (Elsevier, Amsterdam, 1991).
52. Cox, D. L. & Maple, M. B. Electronic pairing in exotic superconductors. *Phys. Today* **40**, 32–40 (1995).
53. Norman, M. R. Hund rule theory for heavy-fermion superconductors. *Phys. Rev. Lett.* **72**, 2077–2080 (1994).
54. Jerome, D. *Organic Conductors* (Dekker, New York, 1994).
55. Ishiguro, T. & Yamaji, K. in *Organic Conductors* Ch. 3 (Springer, Berlin, 1990).
56. Fukuyama, H. in *The Review of High Pressure Science and Technology* (in the press).
57. Levin, R. & Valls, O. T. Strong-coupling theory of superfluid transition temperatures for paramagnon models: application to ³He. *Phys. Rev. B* **17**, 191–200 (1978).
58. Pines, D. Spin excitation and superconductivity in cuprate oxide and heavy electron superconductors. *Physica B* **163**, 78–88 (1990).
59. Moriya, T., Takahashi, Y. & Ueda, K. Antiferromagnetic spin fluctuations and superconductivity in 2-dimensional metals—a possible model for high-T_c oxides. *J. Phys. Soc. Jpn.* **52**, 2905–2915 (1990).
60. Monthoux, P., Balatsky, A. V. & Pines, D. Towards a theory of high temperature superconductivity in the antiferromagnetically correlated cuprate oxide. *Phys. Rev. Lett.* **67**, 3448–3451 (1991).
61. Bulut, N., Hone, D. W., Scalapino, D. J. & Bickers, N. E. Knight-shifts and nuclear-spin-relaxation rates for 2-dimensional models of CuO₂. *Phys. Rev. B* **41**, 1797–1811 (1990).
62. Zhang, S. C. A unified theory based on SO(5) symmetry of superconductivity and antiferromagnetism. *Science* **275**, 1089–1096 (1997).
63. Schröder, A., Aeppli, G., Bucher, E., Ramazashvili, R. & Coleman, P. Scaling of magnetic fluctuations near a quantum phase transition. *Phys. Rev. Lett.* (in the press).
64. von Löhneysen, H. *et al.* Non-Fermi-liquid behaviour in a heavy fermion alloy at a magnetic instability. *Phys. Rev. Lett.* **72**, 3262–3265 (1994).
65. Hirsch, J. E. Attractive interaction and pairing in fermion systems with strong on-site repulsion. *Phys. Rev. Lett.* **54**, 1317–1320 (1985).
66. Miyake, K., Schmitt-Rink, S. & Varma, C. M. Spin-fluctuation mediated even-parity pairing in heavy-fermion superconductors. *Phys. Rev. B* **34**, 6554–6556 (1986).
67. Scalapino, D. J., Loh, E. Jr & Hirsch, J. E. d-wave pairing near a spin-density-wave instability. *Phys. Rev. B* **34**, 8190–8192 (1986).
68. Schrieffer, J. R., Wen, X. G. & Zhang, S. C. Spin-bag mechanism of high-temperature superconductivity. *Phys. Rev. Lett.* **60**, 944–947 (1988).
69. Furukawa, N. & Rice, T. M. Instability of a Landau-Fermi liquid as the Mott insulator is approached. *J. Phys.: Condens. Matter* **10**, L281–L289 (1998).

Acknowledgements. We thank S. V. Brown, P. Agarwal, F. V. Carter, P. Coleman, J. Flouquet, C. D. Frost, I. Gray, D. Khmelntskii, S. J. S. Lister, P. B. Littlewood, A. P. Mackenzie, G. J. McMullan, A. J. Millis, P. Monthoux, C. Pfeleiderer, E. Pugh, S. S. Saxena, A. J. Schofield, M. J. Steiner and A. Tsvetlik. This research has been supported partly by the Cambridge Research Centre in Superconductivity, headed by Y. Liang, by the EPSRC of the UK, by the EU, and by the Cambridge Newton Trust.

Correspondence should be addressed to G.G.L.

Heavy-element enrichment in low-density regions of the intergalactic medium

Lennox L. Cowie & Antoinette Songaila

Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

Models for the composition of the diffuse intergalactic medium^{1,2} predict that low-density intergalactic gas at high redshift should be very poor in heavy elements. This is because locations of early star formation (and thus of heavy-element synthesis) and of gas delivery from such stars are located preferentially within higher-density regions of the intergalactic gas. Here we present a method for analysing carbon and oxygen absorption lines in quasar spectra that allows us to probe the heavy-element abundances at a redshift of three within low-density regions of intergalactic gas. We find that the ratio of triply ionized carbon to neutral hydrogen is roughly constant over a wide range of densities, and that, even as the density approaches zero, the ratio remains high. This unexpected enrichment of low-density gas in heavy elements suggests that early generations of small galaxies might be much more efficient at ejecting heavy elements into the intergalactic medium than has previously been thought.

Recent numerical hydrodynamic simulations^{3–7} have provided detailed descriptions of the gravitational growth of filamentary structures in the intergalactic gas—the ‘cosmic web’—whose fluctuating density results in the absorption lines seen in the spectra of quasars. Neutral hydrogen provides the strongest signal, generating the rich absorption spectrum known as the Lyman- α ‘forest’ at

wavelengths $\lambda < 1,215.67(1 + z_{\text{em}})$ Å, where z_{em} is the quasar’s redshift. However, recent observations^{8,9} using 10-m telescopes have provided spectacular quasar spectra that are surprising in that, even in this relatively diffuse medium, absorption lines from heavy elements are also seen regularly, at least in the higher density portions of the gas. These heavy elements are now thought to have been produced by early generations of stars and dispersed into the more general gas by supernova- or merger-driven winds. They provide, therefore, a powerful probe of the earliest stages in the growth of structure and the formation of galaxies, and it is important to understand how pervasively they are distributed and whether they are present even in the lowest-density regions of the gas.

The transmission in the Lyman- α forest is thought to be an approximate map of the density structure in the low-density intergalactic medium: in particular, the optical depth of neutral hydrogen should roughly determine^{3,5,10–12} the baryon overdensity, $\delta \equiv (\rho_b/\rho_b) - 1$, where ρ_b is the density of baryons. The temperature-density relation¹³ in the low-density intergalactic medium (IGM) has the approximate form $T \sim (1 + \delta)^{0.6}$ as a result of photoionization heating followed by adiabatic cooling. As the total hydrogen density is proportional to $(1 + \delta)$ and the neutral fraction to $(1 + \delta)^{-0.76}$, then $\tau(\text{Ly}\alpha) \propto (1 + \delta)^{1.5}$, where $\tau(\text{Ly}\alpha)$ is the optical depth of the Lyman- α line. Peculiar velocities spread the relationship slightly when it is translated from real to redshift space, but the form of the distribution of τ with δ is well described by the above relationship^{5,12} up to a high overdensity (≥ 10) where shock heating becomes important. The normalization is set by a single parameter, the ratio of the neutral hydrogen ionization rate to the mean baryonic density, and is best obtained by matching the mean Lyman- α opacity in the forest. At redshift $z = 3$, a Lyman- α optical depth of 1 corresponds⁵ roughly to a region in the intergalactic gas that is just slightly overdense ($\delta \approx 0.6$).

Measurement of the abundance of heavy elements in the low-density IGM has traditionally been carried out by identifying strong Ly α absorbing regions (the forest ‘clouds’) and then identifying and measuring metal lines at the same redshift. After pioneering early attempts¹⁴, comprehensive studies^{8,9,15} made possible by the HIRES spectrograph on the Keck I 10-m telescope found a substantial degree of metal enrichment, $[C/H] \approx -2.5$, for forest clouds with total neutral hydrogen column densities, $N(\text{H I}) > 3 \times 10^{14} \text{ cm}^{-2}$ (roughly $\delta \approx 5$ at $z = 3$; see, for example, ref. 1). However, there has been speculation that metallicities may fall rapidly at slightly lower column densities, or equivalently lower overdensities, as galactic enrichment, and the merger expulsion and supernova-generated wind mechanisms responsible for distributing the heavy elements back into the intergalactic gas, should be more efficient in higher overdensity regions¹⁶. In particular, early pre-enrichment models^{1,2} in which the metals are formed in sub-galactic clumps at $z > 5$ predict that at $z \approx 3$ the metallicity should begin to fall rapidly below an overdensity of about $\delta \approx 10$, close to the lowest column densities at which metallicity measurements have been made. This drop in metallicity is predicted¹ to be extremely steep, falling by more than two orders of magnitude from $\delta = 10$ to $\delta = 1$. There is therefore a strong motivation to perform metal determinations at as low an overdensity as possible to investigate whether this prediction is correct.

Previous attempts^{17–19} to measure metallicity at lower densities have generally focused on co-adding metal lines in some way to improve the signal-to-noise ratio. We have developed a very different technique of correlating metal line and H I optical depths. This procedure has many advantages: it avoids subjective Voigt profile fitting, it is very highly sensitive, and it is susceptible to objective assessment of noise levels including systematic uncertainties such as continuum fitting, residual noise from the extraction, and, most importantly, contamination by random lines from systems at other redshifts. This last feature means we have a

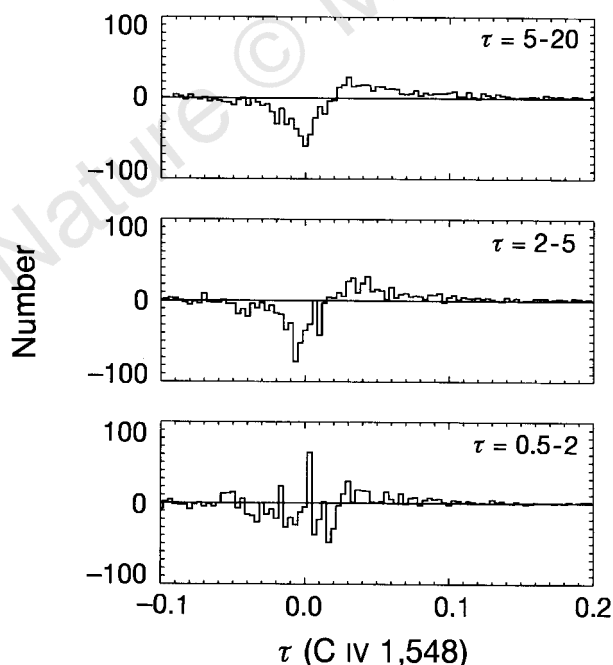


Figure 1 Plots showing the difference between the observed distribution of $\tau(\text{C IV})$ computed in three ranges of $\tau(\text{Ly}\alpha)$, that is $[0.5, 2]$, $[2, 5]$ and $[5, 20]$, and the normalized distribution of $\tau(\text{C IV})$ computed for $\tau(\text{Ly}\alpha) < 0.1$. An excess of positive values is seen in all three ranges. The data set used consisted of observations of six quasars (Q0014 + 813, Q0302 – 003, Q0636 + 680, Q0956 + 122, Q1159 + 123 and Q1422 + 231) obtained with the HIRES spectrograph on the Keck I 10-m telescope. The observations span a redshift range of $z = 2.8$ – 3.4 .

Table 1 Mean and median C IV and O VI opacities

$\tau(\text{Ly}\alpha)$ range	$\langle\tau(\text{Ly}\alpha)\rangle$	$1 + \delta^*$	$\langle\tau(\text{C IV})\rangle^\dagger$	Median $\tau(\text{C IV})^\dagger$	Median $\tau(\text{O VI})^\dagger$
0.5–1	0.71	1.3	0.000 ± 0.002	0.0003 ± 0.0005	0.006 ± 0.004
1–2	1.40	2.0	0.006 ± 0.002	0.0012 ± 0.0004	0.005 ± 0.005
2–5	3.01	3.3	0.012 ± 0.002	0.0033 ± 0.0004	0.007 ± 0.006
5–20	11.55	7.6	0.042 ± 0.003	0.0073 ± 0.0006	0.023 ± 0.007

* δ is the cosmological overdensity, $\delta \equiv (\rho_b/\rho_c) - 1$.

† Formal errors were derived from the $\tau(\text{Ly}\alpha) < 0.1$ sample (Fig. 2 legend).

powerful new ability to analyse metal lines lying in the Lyman- α forest itself.

We first determine $\tau(\text{Ly}\alpha)$ as a function of redshift in each quasar directly for weak absorption ($\tau(\text{Ly}\alpha) \lesssim 3$) in the region between the quasar's Ly β and Ly α emission lines. A few per cent of positions will be contaminated by metal lines²⁰ but this results in a negligible underestimate of the mean and median C IV/H I values that can safely be ignored. The method is also insensitive to the continuum fitting errors in the forest, which correspond to systematic offsets in $\tau(\text{Ly}\alpha)$ that certainly do not exceed 10% at $\tau \approx 1$. Stronger Ly α lines have too little residual flux for their optical depths to be securely determined. In these cases we must trace down through the Lyman series to measure the optical depth and, because higher-order series lines may be contaminated by Lyman- α lines, each subsequent measurement is only an upper limit. For these stronger lines we therefore use $\tau(\text{Ly}\alpha) = \min(\tau(\text{Ly}n) f_{\text{Ly}\alpha} \delta_{\text{Ly}\alpha} / f_{\text{Ly}n} \delta_{\text{Ly}n})$, where the minimum is over all observed lines in the series for which $\tau(\text{Ly}n) > 0.1$ to avoid noise features, and $\tau(\text{Ly}n) < 2$ so that the optical depth can be determined accurately. Here $f_{\text{Ly}n}$ is the oscillator strength of the n th Lyman line (Ly n) and $\delta_{\text{Ly}n}$ is its wavelength. This procedure may overestimate $\tau(\text{Ly}\alpha)$ and so will underestimate metallicity at a given τ but again, in practice, this effect is small. We adopt a similar procedure to minimize contamination in the metal line doublets, setting $\tau(\text{low}) = \min(\tau(\text{low}), 2\tau(\text{high}))$ when $\tau(\text{high})$, the optical depth in the longer wavelength member of the doublet, is greater than 0.05 and otherwise using $\tau(\text{low})$, corresponding to the shorter-wavelength line.

The data base to which we apply the technique is a sample of six $z > 3$ quasars for which complete high signal-to-noise spectra were available from just below 4,000 Å to just above the quasars' C IV

emission lines. The useful redshift range was 2.6–3.4. A description of the data reduction and continuum fitting, along with sample spectra, can be found in ref. 21, where distribution functions of the optical depths in various metal lines can also be found. It is possible to trace down the complete Lyman series only at the highest redshifts (> 3.35), and for these, all points with $\tau(\text{Ly}\alpha) < 200$ were included. Elsewhere, the analysis is restricted to weaker features using optical depths determined from Ly α and Ly β lines with $\tau(\text{Ly}\alpha) > 20$. The corresponding doublets of C IV (1,548 Å) and O VI (1,031 Å) were then measured.

The $\tau(\text{Ly}\alpha) < 20$ sample contains just under 70,000 points (these are not all statistically independent because of velocity broadening and instrumental resolution) of which approximately half have $\tau(\text{Ly}\alpha) < 0.1$. We used this latter distribution as a representative sample of random points to determine backgrounds and errors and to measure the significance of detections (Fig. 2 legend). As our first step in the analysis we formed the distributions of $\tau(\text{C IV})$ in three $\tau(\text{Ly}\alpha)$ intervals, namely [0.5, 2], [2, 5] and [5, 20], and compared these with distributions from the reference 'blank' sample with $\tau(\text{Ly}\alpha) < 0.1$. The differences between the three distributions and the normalized blank distribution is shown in Fig. 1. In each case a positive excess is seen as an 'S' wave in the plot, though for $\tau(\text{Ly}\alpha) = 0.5$ –2 the signal is becoming weak. At lower $\tau(\text{Ly}\alpha)$ the signal can no longer be seen.

We find significant signal in both the median and mean C IV opacities down to $\tau(\text{Ly}\alpha) = 1$ (summarized in Table 1). To show this in more detail we have divided the sample, ordered by $\tau(\text{Ly}\alpha)$, into 1,500-element bins and computed the median and mean in each bin. Once again we have computed the error based on the dispersion in the bins at $\tau(\text{Ly}\alpha) < 0.1$. The

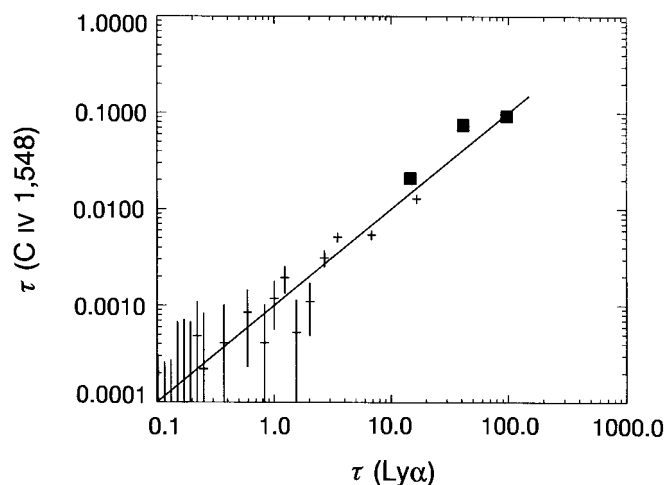


Figure 2 The median $\tau(\text{C IV})$ in 1,500-element bins (crosses) as a function of $\tau(\text{Ly}\alpha)$. The error bars are $\pm 1\sigma$ based on the dispersion in the $\tau(\text{Ly}\alpha) < 0.1$ sample, determined by dividing the blank sample into a large number of subsamples each with numbers equal to those in the various bins, and then computing the dispersions of the errors in medians and means. The filled squares show the higher- τ sample from the more restricted redshift range ($z = 3.34$ – 3.41 in Q1159 + 123 and Q1422 + 231) for which full Lyman series coverage is available. The solid line is the relationship $\tau(\text{C IV}) = 10^{-3} \tau(\text{Ly}\alpha)$. We have checked the

methodology with a number of tests to make sure there are no systematic effects. Most importantly, displacing the metal lines by $> 100 \text{ km s}^{-1}$ from their positions eliminates the signal, whereas eliminating all points lying with 200 km s^{-1} of any position with $\tau(\text{Ly}\alpha) < 14$ does not change the significance of the signal at $\tau = 1$ –2. The latter result confirms that the positive signal does not arise primarily in the outlying regions of higher column density zones, but rather is more uniformly spread through the entire volume.

results are shown in Fig. 2. They have been extended to higher τ using the $\tau(\text{Ly}\alpha) < 200$ sample divided into 50-point bins. Again we find that there is a significant signal in both the median and the mean down to $\tau(\text{Ly}\alpha) = 1$, corresponding to only very slightly over-dense gas. The least-square power-law fit to the median $\tau(\text{CIV})$ versus $\tau(\text{Ly}\alpha)$ relation between $\tau(\text{Ly}\alpha) = 1$ and $\tau(\text{Ly}\alpha) = 20$ gives $\tau(\text{CIV}) = 0.0008 \tau(\text{Ly}\alpha)^{1.03}$, whereas a similar fit to the mean gives $\tau(\text{CIV}) = 0.0047 \tau(\text{Ly}\alpha)^{0.90}$. Both are consistent with a constant value of $\text{CIV}/\text{H I}$ as a function of $\tau(\text{Ly}\alpha)$ which is similar to the values measured at higher H I column densities. In the $2 < \tau(\text{Ly}\alpha) < 5$ range, $N(\text{CIV})/N(\text{H I}) = f_{\text{Ly}\alpha} \lambda_{\text{Ly}\alpha} \tau(\text{CIV}) / f_{\text{CIV}} \lambda_{\text{CIV}} \tau(\text{Ly}\alpha)$ has a mean of $(7.0 \pm 2.3) \times 10^{-3}$ and a median of $(1.9 \pm 0.7) \times 10^{-3}$, compared to the mean of 3.5×10^{-3} and median of 3×10^{-3} determined for clouds in the column density range $N(\text{H I}) = 10^{15} - 10^{17} \text{ cm}^{-2}$ and the median of 2×10^{-3} for $N(\text{H I}) = (5 \times 10^{14}) - 10^{15} \text{ cm}^{-2}$ given in ref. 9. In the present method, the median may be skewed systematically to lower values because thermal broadening spreads the H I lines over a larger wavelength range than the CIV , and the mean may represent the best estimate of the absolute value, though there is clearly uncertainty in the absolute $\text{CIV}/\text{H I}$ determination at the factor 2–3 level. However, irrespective of this point, the results suggest that the average value of $\text{CIV}/\text{H I}$ is roughly constant over a wide range of $N(\text{H I})$ extending from $\sim 3 \times 10^{13} \text{ cm}^{-2}$ to near the level of $N(\text{H I}) \geq 10^{17} \text{ cm}^{-2}$ at which radiative transfer effects become important.

We note that work recently reported²² by Lu *et al.* uses a version of the stacking technique to infer a much lower value ($< 3.89 \times 10^{-4}$) of $\text{CIV}/\text{H I}$ at low $N(\text{H I})$ than is derived here. Interestingly, Lu¹⁸ used a very similar version of this method in 1991 to analyse pre-Keck quality spectra in the range $10^{15} < N(\text{H I}) < 10^{17} \text{ cm}^{-2}$ for which direct measurements of $\text{CIV}/\text{H I}$ can now be obtained⁹. His derived limit ($\text{CIV}/\text{H I} < 2 \times 10^{-4}$) is also almost an order of magnitude less than the actual measured mean or median values, which lie between 2×10^{-3} and 3.5×10^{-3} at these column densities. The reason for this systematic difference is not fully clear, but it may arise from the methods of handling the $N(\text{H I})$ distributions and from the direct culling of higher $N(\text{CIV})$ systems in the particular stacking method used. We shall provide elsewhere a more detailed critical assessment of various stacking techniques.

The other metal line that is expected^{15,23} to be strong in the weak forest is the O VI ($\lambda 1,032 \text{ \AA}$) doublet of five times ionized oxygen. However, the exact $\text{O VI}/\text{H I}$ value is very dependent on the shape of the high-energy ionization field¹⁵. The O VI line lies deep in the $\text{Ly}\alpha$ and $\text{Ly}\beta$ forest, making conventional analysis very difficult, but it is susceptible to analysis by the present technique because in the ensemble of optical depths the forest contamination simply introduces an additional noise term. The median $\tau(\text{O VI})$ is summarized in Table 1. Because of the higher noise in the forest, the O VI is detected only above a value of $\tau(\text{Ly}\alpha) = 5$ at which $\tau_{\text{median}}(\text{O VI}) / \tau_{\text{median}}(\text{CIV})$ is 3.1 ± 0.9 , but we can place a 1σ upper limit of $\tau_{\text{median}}(\text{O VI}) / \tau_{\text{median}}(\text{CIV}) \leq 8$ at $1 < \tau < 2$. The value of $\text{O VI}/\text{H I}$ is 8.8×10^{-3} at $\tau = 5 - 20$.

Translation of $\text{CIV}/\text{H I}$ into C/H is uncertain because it depends on the poorly known shape of the photoionizing spectrum, and may also be confused by collisional ionization contributions²⁴. If the metallicity were invariant with density, the relative constancy of $\text{CIV}/\text{H I}$ as the optical depth falls is unexpected if the background ionizing field is a pure power law or has a spectrum which more accurately represents the sum of the quasars and the radiative transfer through the IGM²⁵; this is because the decreasing density would imply an increasing ionization parameter, with the result that a large fraction of the carbon would appear as C V and the value of $\text{CIV}/\text{H I}$ should decline. In order to maintain a fairly constant value of $\text{CIV}/\text{H I}$ we require a spectrum that is heavily broken across the He^+ edge at 54 eV, which then forces the carbon to remain in the C IV

level¹⁵. This result is consistent with interpretations⁹ of the behaviour of $\text{Si IV}/\text{C IV}$ at these redshifts. The values of $\text{O VI}/\text{C IV}$ are then approximately consistent with the observed values¹⁵.

More importantly, our results require that there be a substantial metallicity at only very slight overdensities in the IGM. Even in the case of the heavily broken power law, which maximizes the value of $\text{CIV}/\text{H I}$ at low densities, the maximum value of $\text{CIV}/\text{H I}$ reaches just over 0.1 for C/H of 0.01 times the solar value^{15,23}. The mean value of C/H would then exceed 7×10^{-4} and the median exceed 1.4×10^{-4} throughout the density range, only slightly lower than the average metallicity derived for the IGM²⁶. For more energetic ionizing spectra, abundances one to two orders of magnitude higher would be required^{15,23}. This high level of distributed metallicity suggests that the ejection mechanisms in $z > 5$ subgalactic clumps are more efficient than has been thought, or that there is some alternative, more uniform, source of heavy elements which we do not currently understand. If some part of the metallicity is not attributable to subgalactic-scale galaxy formation, we may be overestimating the number of $z > 5$ type-II supernovae and star-formation regions whose detection is one of the primary justifications^{27,28} for the Next Generation Space Telescope. $\square$

Received 27 January; accepted 21 April 1998.

- Gnedin, N. Y. Metal enrichment of the intergalactic medium. *Mon. Not. R. Astron. Soc.* **294**, 407–421 (1998).
- Gnedin, N. Y. & Ostriker, J. P. Reionization of the universe and the early production of metals. *Astrophys. J.* **486**, 581–598 (1997).
- Cen, R., Miralda-Escudé, J., Ostriker, J. P. & Rauch, M. Gravitational collapse of small-scale structure as the origin of the Lyman-alpha forest. *Astrophys. J.* **437**, L9–L12 (1994).
- Petitjean, P., Mückert, J. P. & Kates, R. E. The Ly α forest at low redshift: tracing the dark matter filaments. *Astron. Astrophys.* **295**, L9–L12 (1995).
- Zhang, Y., Anninos, P. & Norman, M. L. A multispecies model for hydrogen and helium absorbers in Lyman-alpha forest clouds. *Astrophys. J.* **453**, L57–L60 (1995).
- Hernquist, L., Katz, N., Weinberg, D. H. & Miralda-Escudé, J. The Lyman-alpha forest in the cold dark matter model. *Astrophys. J.* **457**, L51–L56 (1996).
- Miralda-Escudé, J., Cen, R., Ostriker, J. P. & Rauch, M. The Ly alpha forest from gravitational collapse in the cold dark matter + lambda model. *Astrophys. J.* **471**, 582–616 (1996).
- Cowie, L. L., Songaila, A., Kim, T.-S. & Hu, E. M. The metallicity and internal structure of the Lyman alpha forest clouds. *Astron. J.* **109**, 1522–1530 (1995).
- Songaila, A. & Cowie, L. L. Metal enrichment and ionization balance in the Lyman alpha forest at $z \sim 3$. *Astron. J.* **112**, 335–351 (1996).
- Bi, H. G. & Davidson, A. Evolution of structure in the intergalactic medium and the nature of the Ly alpha forest. *Astrophys. J.* **479**, 523–542 (1997).
- Davé, R., Hernquist, L., Weinberg, D. H. & Katz, N. Voigt-profile analysis of the Ly alpha forest in a cold dark matter universe. *Astrophys. J.* **477**, 21–26 (1997).
- Croft, R. A. C., Weinberg, D. H., Katz, N. & Hernquist, L. Intergalactic helium absorption in cold dark matter models. *Astrophys. J.* **488**, 532–549 (1997).
- Hui, L. & Gnedin, N. Y. Equation of state of the photoionized intergalactic medium. *Mon. Not. R. Astron. Soc.* **292**, 27–42 (1997).
- Meyer, D. & York, D. G. Observations of weak C IV absorption toward the QSOs 2000-330 and 2126-158. *Astrophys. J.* **315**, L5–L9 (1987).
- Rauch, M., Haehnelt, M. G. & Steinmetz, M. QSO metal absorption systems at high redshift and the signature of hierarchical galaxy formation. *Astrophys. J.* **481**, 601–624 (1997).
- Petitjean, P. in *Science with the VLT* (eds Danziger, J. & Walsh, J.) 339 (Springer, Heidelberg, 1995).
- Norris, J., Peterson, B. A. & Hartwick, F. D. A. The oxygen abundance in the clouds producing the Ly-alpha-Ly-beta absorption systems in the quasars 4C 5.34 (0.804+0.046) and OQ 172 (1442+101). *Astrophys. J.* **273**, 450–457 (1983).
- Lu, L. The carbon abundance of the Lyman-alpha clouds. *Astrophys. J.* **379**, 99–106 (1991).
- Tytler, D. & Fan, X.-M. Upper limits on metals in quasar Lyman-alpha forest clouds: Absence of C IV lines in echelle spectra. *Astrophys. J.* **424**, L87–L90 (1994).
- Hu, E. M., Kim, T.-S., Cowie, L. L. & Songaila, A. The distribution of column densities and b-values in the Lyman-alpha forest. *Astron. J.* **110**, 1526–1543 (1995).
- Songaila, A. The redshift evolution of the metagalactic ionizing flux inferred from metal line ratios in the Lyman forest. *Astron. J.* (in the press).
- Lu, L., Sargent, W. L. W., Barlow, T. A. & Rauch, M. The metal contents of very low column density Lyman-alpha clouds: implications for the origin of heavy elements in the intergalactic medium. Preprint astro-ph/9802189.
- Hellstein, U., Davé, R., Hernquist, L., Weinberg, D. H. & Katz, N. Metal lines associated with Ly alpha absorbers: A comparison of theory and observations. *Astrophys. J.* **487**, 482–488 (1997).
- Giroux, M. L. & Shull, J. M. The influence of the photoionizing radiation spectrum on metal-line ratios in Ly (alpha) forest clouds. *Astron. J.* **113**, 1505–1513 (1997).
- Haardt, F. & Madau, P. Radiative transfer in a clumpy universe. II. The ultraviolet extragalactic background. *Astrophys. J.* **461**, 20–37 (1996).
- Songaila, A. A lower limit to the universal density of metals at $z \sim 3$. *Astrophys. J.* **490**, L1–L4 (1997).
- Haiman, Z. & Loeb, A. Signatures of stellar reionization of the universe. *Astrophys. J.* **483**, 21–37 (1997).
- Miralda-Escudé, J. & Rees, M. J. High-redshift supernovae and the metal-poor halo stars: Signatures of the first generation of galaxies. *Astrophys. J.* **478**, L57–L61 (1997).

Acknowledgements. We thank the many people at the Keck telescopes who assisted with these observations, particularly S. Vogt whose HIRES spectrograph made the observations possible. L.L.C. and A.S. are visiting astronomers at the W. M. Keck Observatory, jointly operated by the California Institute of Technology and the University of California. This work was supported by the US NSF.

Correspondence and requests for materials should be addressed to L.L.C. (e-mail: cowie@uhifa.hawaii.edu).

Quantum error correction for communication with linear optics

Samuel L. Braunstein

SEECs, University of Wales, Bangor LL57 1UT, UK

Improving the signal-to-noise ratio in optical communication systems is a fundamental requirement for cost-effective data transmission. This is particularly important for the transmission of noise-intolerant quantum states: excess noise at the quantum level destroys the coherence of the states, rendering classical error correction or amplifier-based schemes¹ useless for quantum communication. Only quantum error correction^{2,3} can remove the effects of noise without corrupting the fragile superpositions of quantum states. But difficulties arise in the practical implementation of such a correction process because nonlinear operations⁴ have been thought to be required, greatly reducing the efficiency of any optical scheme. Here I report an efficient, compact scheme involving only linear optical elements and feedback, which performs error correction for both quantum and classical noise. In the classical case, the noise penalty incurred is no worse than for ideal amplification. But for low-noise quantum optical communication, this penalty may be eliminated entirely. This quantum error-correction scheme may thus find application in quantum cryptographic networks⁵⁻⁷ (where low noise is equivalent to high security), possibly extending their range far beyond limits imposed by system losses⁷.

The discovery of quantum error-correcting codes^{2,3} substantially enhanced the feasibility of building a quantum computer and of one day using it to efficiently factor large numbers⁸. The basis of this substantial advance lies in the ability to tame the hazards of decoherence by encoding the information of a given state in a set of entangled states, without violating the no-cloning theorem of quantum mechanics⁹. Recently, quantum error-correction protocols have been extended from discrete quantum variables, such as quantum bits (qubits), to continuous ones, such as arbitrary-intensity optical wavepackets^{10,11}. Implementing continuous quantum error correction in optical communications systems is, therefore, the next logical step. Yet, the prevailing view had been that nonlinear operations, such as the exclusive-OR gate⁴, are essential for a compact construction. According to this view, replacing these nonlinear components by linear ones would require an exponential increase in their number¹². A hint that we may be able to use only a handful of linear optical elements when dealing with continuous variables is found in a recent realization¹³ of quantum teleportation using only beam-splitters, feedback and a source of optical squeezed states¹⁴. The degree of squeezing would determine the noise penalty in the correction fidelity: completely classical light would cause relatively small, finite errors, and perfect squeezing would yield an ideal quantum optical repeater.

The implementation I describe here involves a nine-wavepacket code, which completely encodes the full quantum state of a single wavepacket such that if one of those nine wavepackets were to be disturbed in any way the original information may be recovered^{10,11}. This nine-wavepacket code is the continuous-variable analogue of Shor's original code for discrete quantum states². To implement Shor's qubit code using only linear operations would require $2^9 = 512$ channels and exponentially many more components¹². By contrast, the code described here uses nine channels and encoding is performed with only eight beam-splitters. An optimal-efficiency five-wavepacket code (with the same encoding and noise-resistant properties) has recently been found¹⁰. I prefer to

work with the nine-wavepacket code here, owing to its simple and intuitive structural properties.

I focus attention on the state of a single polarization of a transverse mode of electromagnetic radiation, which may be succinctly described by a one-dimensional wavepacket. As the electric quadrature amplitudes (the real and imaginary parts of the complex electric field) form a canonically conjugate pair, in analogy with position and momentum, it is convenient to use the latter terms. In a units-free notation:

$$\text{position} = xL$$

$$\text{momentum} = p/L \quad (1)$$

where x and p are scaled position and momentum (with some arbitrary length scale L) and $\hbar = 1/2$. (I henceforth drop the modifier 'scaled'.) The position basis eigenstates $|x\rangle$ are normalized according to $\langle x'|x\rangle = \delta(x' - x)$ with the momentum basis given by:

$$|p\rangle = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dx e^{2ixp} |x\rangle \quad (2)$$

To avoid confusion, I shall work in the position basis throughout. The Fourier transform is defined as an active operation on a state by

$$\hat{F}|x\rangle = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dy e^{2ixy} |y\rangle \quad (3)$$

where both x and y are variables in the position basis and symbols marked with a circumflex denote linear operators. Note that equations (2) and (3) correspond to a change of representation and a physical change of state, respectively. For optical wavepackets this Fourier transform describes the action of a $1/4$ -wave delay.

It turns out that all one needs to construct an encoding element is a series of beam-splitters and phase delays. The linearity of the coding device is facilitated by the freedom in the choice of length scale in equation (1), which is meaningful only for systems with continuous variables.

I now seek a device that will encode a position-state eigenstate $|x\rangle$ in a nine-wavepacket code of the form¹⁰:

$$|x_{\text{encode}}\rangle = \frac{1}{\pi^{3/2}} \int dw dy dz e^{2ix(w+y+z)} \times |w, w, w, y, y, y, z, z, z\rangle \quad (4)$$

First, consider two light fields $|x\rangle$ and $|y\rangle$ incident on an 'ideal' (phase-free) beam-splitter. The output light fields are given by:

$$\hat{B}_{12}(\theta)|x, y\rangle = |x \cos \theta - y \sin \theta, y \cos \theta + x \sin \theta\rangle \quad (5)$$

where the subscripts 12 refer to the wavepackets acted upon. From these ideal beam-splitters I construct a three-port device called a tritter¹⁵:

$$\hat{T}_{123} \equiv \hat{B}_{23}(\pi/4) \hat{B}_{12} \cos^{-1} \frac{1}{\sqrt{3}} \left(\right) \quad (6)$$

for which $\hat{T}_{123}|x, 0, 0\rangle = |x/\sqrt{3}, x/\sqrt{3}, x/\sqrt{3}\rangle$. Finally, all the above elements are combined to produce the encoding device. A suitable choice for the nine-wavepacket code is a nine-port beam-splitter, which I shall call a nona-splitter:

$$\hat{N}_{1-9} \equiv \hat{T}_{789} \hat{T}_{456} \hat{T}_{123} \hat{F}_7 \hat{F}_4 \hat{F}_1 \hat{T}_{147} \quad (7)$$

where I have not explicitly included the free propagation factors between optical elements. This device could be implemented either as a series of eight ordinary beam-splitters or as a single (mass-produced) integrated-optics element.

The nine-wavepacket code, equation (4), may now be formed with the nona-splitter acting on the initial unencoded state, via $|x_{\text{encode}}\rangle \propto \hat{N}_{1-9}|x_{\text{init}}\rangle$. Written in this order, $\hat{F}_7 \hat{F}_4 \hat{F}_1 \hat{T}_{147}$ forms the momentum-error-correction code (or Fourier-transformed position code), and the remaining tritters $\hat{T}_{789} \hat{T}_{456} \hat{T}_{123}$ form the position-error-correction codes for each of the three momentum-encoded wavepackets.

By linearity, it is sufficient to consider the encoding procedure on

a position eigenstate $|x\rangle$. The eight auxiliary inputs are prepared with zero-position eigenstates (that is, ideally squeezed states, an assumption that is relaxed below), so the initial unencoded state has the form $|x_{\text{ini}}\rangle = |x, 0, 0, 0, 0, 0, 0, 0\rangle$. Having introduced the coding device, I proceed with a discussion of the error detection and correction procedure. The three states constituting error correction described here are: inverted encoding, error-syndrome identification and error correction. In the case of perfect signal-transmission, inverted encoding trivially recovers the initial (unencoded) state with the auxiliary wavepackets returning to their zero positions. If, however, one of the encoded wavepackets suffers from noise, then following inverted encoding the auxiliary wavepackets will display a non-zero 'syndrome', indicating the type and size of the error. More specifically, position measurements are performed on each of the eight auxiliary ports. Syndrome identification then involves only simple comparisons between these real numbers. Once identified, the syndrome is translated to the appropriate remedy which, for my scheme, corresponds to a pure linear displacement (whose parameters are obtained by a linear combination of these real numbers). This remarkable fact, that an arbitrary error can be corrected with a linear displacement, is general and very important for purposes of implementation.

As an example, consider an inverted encoding algorithm which inverts $\hat{N}_{1-9}$ in two steps, enabling independent correction of position and momentum errors. First I invert the last three tritters (those to the left) in equation (7) and identify the syndrome for the position error. There are three sets of syndromes here corresponding to the positions of auxiliary wavepackets (2, 3), (5, 6) and (8, 9). Then I invert the remaining terms to obtain the momentum error. The syndrome in this case is given by the positions of auxiliary wavepackets 4 and 7. Finally, I perform a linear displacement based on the identified syndrome and retrieve the original wavepacket.

So far, I have presented the error-correction code and procedure. This code will operate ideally (that is, perform perfect error-correction) under noise-free conditions, corresponding to ideal-squeezing of the auxiliary wavepackets, noiseless manipulation and perfect detection. (Indeed, this is also the prediction one would obtain from classical wave theory.) What would be the penalty of operating this scheme with imperfect resources and detection? As a first source of errors, I examine the use of auxiliary wavepackets with an initial squeezing parameter r and uncertainty $e^{-r}/2$. Moreover, I suppose that the position measurements necessary for reading the error-syndrome are carried out by homodyne detectors with an energy efficiency η , yielding a noise term that is represented by vacuum leakage (with uncertainty $1/2$) into the device¹³. In this case, the j th component of the syndrome x_{syn} (that is, the position measured at the j th output port) becomes randomized by the uncertainty. Thus, for an error originating on the k th wavepacket ($k = 1 \dots 9$) each component of the syndrome takes the generic form:

$$x_{\text{syn}}^j = x_{0,\text{syn}}^j + \sum_i c_{ji}^k x_{\text{sq}}^i + \sqrt{\eta^{-1} - 1} x_{\text{vac}}^j \quad (8)$$

Here $x_{0,\text{syn}}$ is the ideal syndrome, and x_{sq} and x_{vac} are noise terms due to limited squeezing of the input states and imperfect detection at the output, respectively, for some constants c_{ji}^k .

Under these noisy conditions, some non-zero syndrome will always be identified, even in the absence of an actual error (to the encoded state). Correcting fictitious errors accumulates small yet undesired penalties. A more careful, bayesian, approach to syndrome identification may avoid this and would be especially useful when real errors occur rarely per error-correction cycle. Precisely how big would the penalty be if correction was carried through? After displacement, the output state may be represented as a Wigner function by a convolution of the Wigner function of the original unencoded state with an error gaussian $\exp[-2|\alpha|^2/(e^{-2r} + \eta^{-1} - 1)]$ (up to normalization). For the case

of efficient detection but auxiliary wavepackets in the vacuum state, this gaussian reduces to the Wigner representation for vacuum, and a noise penalty of roughly one unit of vacuum fluctuations is incurred¹⁶. By squeezing the auxiliary wavepackets, even this penalty may be reduced to allow for long-distance, error-corrected quantum state transmission.

Perhaps the most promising potential application of quantum error correction within an optical channel lies in quantum cryptography. Today, system losses impose a limit to the range of quantum communication⁷. Quantum error correction can be applied to attack this problem, both in the case of the cryptographic transmission of continuous variables and in more conventional qubit-transmission. In the former case, this calls for further consideration of recently proposed schemes for ideal (noiseless) continuous-variable quantum cryptography^{17,18}. For finite squeezing or imperfect detection, however, some residual noise will be present. This noise has the effect of helping the eavesdropper mask her attack, so our guarantee of absolute security is lost.

The ability to encode qubits in continuous variables suggests a more likely alternative for implementation within my quantum error-correction scheme. There are several ways qubits could be represented within our scheme. They could be encoded as polarizations (requiring a duplication of our scheme for each polarization channel)⁵. Less expensive alternatives might involve encoding qubits as the phase between pulses in unbalanced Mach-Zehnder interferometers⁶, or encoding them as the time between pulses in interferometers using Faraday mirrors⁷. All of these schemes have been implemented in quantum cryptography. Unlike continuous variable schemes, there exist well-developed methods for dealing with noise in qubit-based quantum cryptography including classical¹⁹ and quantum²⁰ privacy amplification.

Finally, I consider the performance of my error-correction scheme as a classical tool for long-distance communication and data-transmission. In particular, can error correction provide a realistic alternative to amplifier-based repeaters? First, I note that amplifier-based repeaters are phase insensitive, while for error-correction a high degree of interferometric stability is required. Having said this, the low noise in my scheme still offers possible advantages over conventional amplification methods. Even though the basic penalty of one unit of vacuum fluctuations for my scheme is equal to that of strongly amplifying a signal with an ideal amplifier¹, distributing amplifiers along a channel introduces extra complications: an amplifier placed at the beginning of a channel produces a greater energy throughput, thus requiring higher channel capacities. By comparison, in error-correction schemes, a fixed reduction of channel capacity is incurred due to the transmission of redundant information in the auxiliary wavepackets. This penalty can, in principle, be made arbitrarily small for asymptotically efficient codes. Amplifiers also suffer penalties when placed at the end of channels, as the additional noise from the channel is amplified along with the signal. By contrast, my scheme is designed to correct errors, rather than maintain the signal-to-noise ratio. One example of a difficulty faced by conventional amplification schemes is the occurrence of large burst errors in space communications. Using only classical resources (that is, $r = 0$), my scheme would correct single wavepacket errors independent of the burst intensity.

Quantum error correction will almost certainly play an important role in new technologies based on quantum information. I have shown that continuous error correction can be achieved in an interferometric set-up with only a handful of linear components and a standard quantum resource: squeezed states of light. By comparison, standard quantum computational networks acting on discrete states require similar numbers of nonlinear components. Maintaining the interferometric stability required for operating the proposed scheme would appear to be feasible with today's technology. Moreover, I have shown that even in the absence of

squeezing, the scheme's performance is remarkably good, closely competing with ideal amplifier-based repeaters. Replacing such repeaters may become the first implementation of quantum-computational concepts to (real-world) classical problems. $\square$

Received 8 December 1997; accepted 14 April 1998.

1. Haus, H. A. & Mullen, J. A. Quantum noise in linear amplifiers. *Phys. Rev.* **128**, 2407–2413 (1962).
2. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, R2493–R2496 (1995).
3. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
4. DiVincenzo, D. P. & Smolin, J. in *Proc. Workshop on Physics and Computation PhysComp '94* 14–23 (IEEE Computer Soc. Press, Los Alamitos, CA, 1994).
5. Bennett, C. H. & Brassard, G. The dawn of a new era for quantum cryptography: the experimental prototype is working! *SIGACT News* **20**, 78–82 (1989).
6. Townsend, P. D., Rarity, J. G. & Tapster, P. R. Enhanced single photon fringe visibility in a 10 km-long prototype quantum cryptography channel. *Electron. Lett.* **29**, 1291–1293 (1993).
7. Muller, A. *et al.* Plug and play systems for quantum cryptography. *Appl. Phys. Lett.* **70**, 793–795 (1997).
8. Shor, P. W. in *Proc. 35th Annual Symp. on the Foundations of Computer Science* (ed. Goldwasser, S.) 124–134 (IEEE Computer Soc. Press, Los Alamitos, CA, 1994).
9. Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).
10. Braunstein, S. L. Error correction for continuous quantum variables. *Phys. Rev. Lett.* **80**, 4084–4087 (1998).
11. Lloyd, S. & Slotine, J.-J. E. Analog quantum error correction. *Phys. Rev. Lett.* **80**, 4088–4091 (1998).
12. Reck, M. *et al.* Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58–61 (1994).
13. Braunstein, S. L. & Kimble, H. J. Teleportation of continuous quantum variables. *Phys. Rev. Lett.* **80**, 869–872 (1998).
14. Walls, D. F. Squeezed states of light. *Nature* **306**, 141–146 (1983).
15. Zukowski, M. Entanglement and photons. *Laser Phys.* **4**, 690–709 (1994).
16. Musslimani, Z. H., Braunstein, S. L., Mann, A. & Revzen, M. Destruction of photocount oscillations by thermal noise. *Phys. Rev. A* **51**, 4967–4973 (1995).
17. Mu, Y., Seberry, J. & Zheng, Y. Shared cryptographic bits via quantized quadrature phase amplitudes of light. *Opt. Commun.* **123**, 344–352 (1996).
18. Cohen, O. Quantum cryptography using nonlocal measurements. *Helv. Phys. Acta* **70**, 710–726 (1997).
19. Bennett, C. H., Brassard, G. & Robert, J.-M. Privacy amplification by public discussion. *SIAM J. Comput.* **17**, 210–229 (1988).
20. Deutsch, D. *et al.* Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Phys. Rev. Lett.* **77**, 2818–2821 (1996).

Acknowledgements. I thank N. Cohen, H. J. Kimble and A. M. Steane for discussions.

Correspondence and requests for materials should be addressed to the author (e-mail: schmuel@sees.bangor.ac.uk).

Nanofabrication of solid-state Fresnel lenses for electron optics

Y. Ito, A. L. Bleloch & L. M. Brown

Department of Physics, University of Cambridge, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

Lenses for precision electron optics are mainly magnetic, requiring large cylinders of soft iron to focus an electron beam. Such lenses can only be convergent¹, and so suffer from spherical aberration. Electrostatic lenses are sometimes used, but tend to be even more cumbersome. The advent of high-brightness electron guns for scanning transmission electron microscopy has made it possible to use the resulting tightly focused electron beams to drill holes a few nanometres in size and of controlled depth in some inorganic thin films^{2–5}: such patterned structures can then be used to manipulate the phase of an electron wave in a manner analogous to light optics^{6,7}. Here we use this approach to fabricate compact solid-state 'pixelated' Fresnel lenses for electron optics. These lenses, which can be convergent or divergent, are not expected to compete with conventional magnetic lenses in most applications (such as microscopy), but may find a niche in electron-beam lithography.

In order to focus an incident electron plane wave (wavelength λ), it must undergo a phase retardation $\phi(r)$ which varies with the radius from the optic axis, r , as follows:

$$\phi(r) = k_0(f - (f^2 + r^2)^{1/2}) \quad (1)$$

where $k_0 = 2\pi/\lambda$ and f is the focal length of the lens⁸.

$\phi(r)$ can be modified to have a modulo 2π phase structure, and the phase distribution of a zone, $\phi_F(r)$, becomes:

$$\phi_F(r) = \phi(r) + 2m\pi, \quad r_m < r < r_{m+1} \quad (2)$$

where r_m is the inner radius of the m th zone. Equation (2) can, of course, be modified to include any additional phase to correct

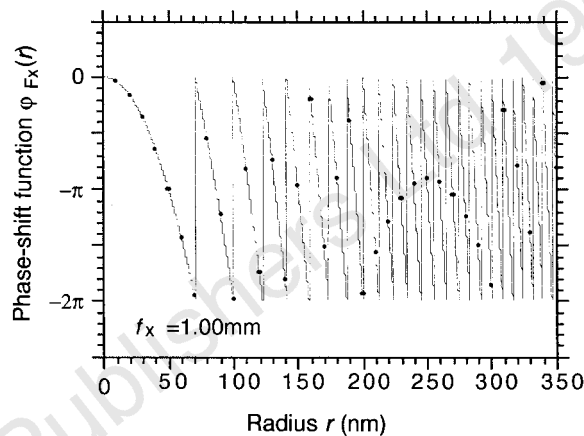


Figure 1 Phase-shift function, $\phi_F(r)$, of the ideal Fresnel lens (line) and the radial positions of the holes and their expected phase shifts for the actual lens (dots) with a nominal focal length of $f = 1$ mm. Because the dwell time of the beam for each hole is controlled, the expected phase values were calculated from the linear relationship between the dwell time and the desired phase here. For the higher-order zones, an aliasing effect between the hole position and $\phi_F(r)$ can be seen. The minimum zone width is limited by the size of a single hole (~ 2 nm).

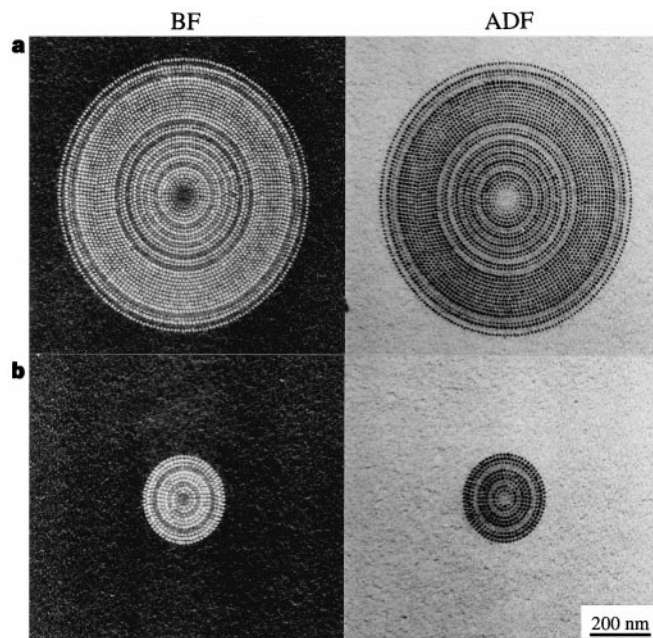


Figure 2 Two Fresnel lenses with different focal lengths. Bright field (BF) and high angle annular dark field (ADF) images of nominal focal length of **a**, $f = 1$ mm and **b**, $f = 0.25$ mm lenses. The diameters of each lens are **a**, 700 nm and **b**, 236 nm. Before observation, the lenses were fabricated in VG HB501 STEM with a probe current of ~ 200 pA at 100 keV. An α -AlF₃ film, 68 nm thick, was overwritten 30 times with the lens pattern. The total fabrication time for the lenses was 60 s and 6 s for **a** and **b**, respectively. A single zone has the same width as the hole spacing from the 10th ($r_m = 316$ nm) and 3rd ($r_m = 112$ nm) zone outwards for **a** and **b**, respectively. The outer zones beyond half of the radius demonstrate this aliasing effect.

aberrations (except chromatic).

Hence,

$$r_m = (2m\lambda f + (m\lambda)^2)^{1/2} \quad (3)$$

$$f = (r_m^2 - (m\lambda)^2)/(2m\lambda) \quad (4)$$

The phase function of the pixelated Fresnel lens (PFP) lens will have equation (2) as an envelope function multiplied by a two-dimensional array of holes as shown in Fig. 1.

In order to control the phase of the electrons it is necessary to drill holes of known depth in a thin film, the inner potential of which provides the refractive power of the lens. Methods for doing this

have been described previously⁶. Briefly, the focused beam used to fabricate the lens is made to dwell at a particular point for a time that is controlled by computer. The gradient of hole depths can be varied by optimizing the sequence of dwell times. Thus it is possible to create a pattern of holes that changes the phase systematically across a circular Fresnel zone. The largest lens described here requires about 1 min to fabricate, is about 0.7 μm in diameter and contains 3,981 holes.

The focal length and the number of zones achievable is strongly influenced by the film thickness, minimum hole spacing and achievable maximum pattern size. For 200 keV electrons, an amorphous AlF_3 film 72 nm thick gives a phase change close to 2π at the

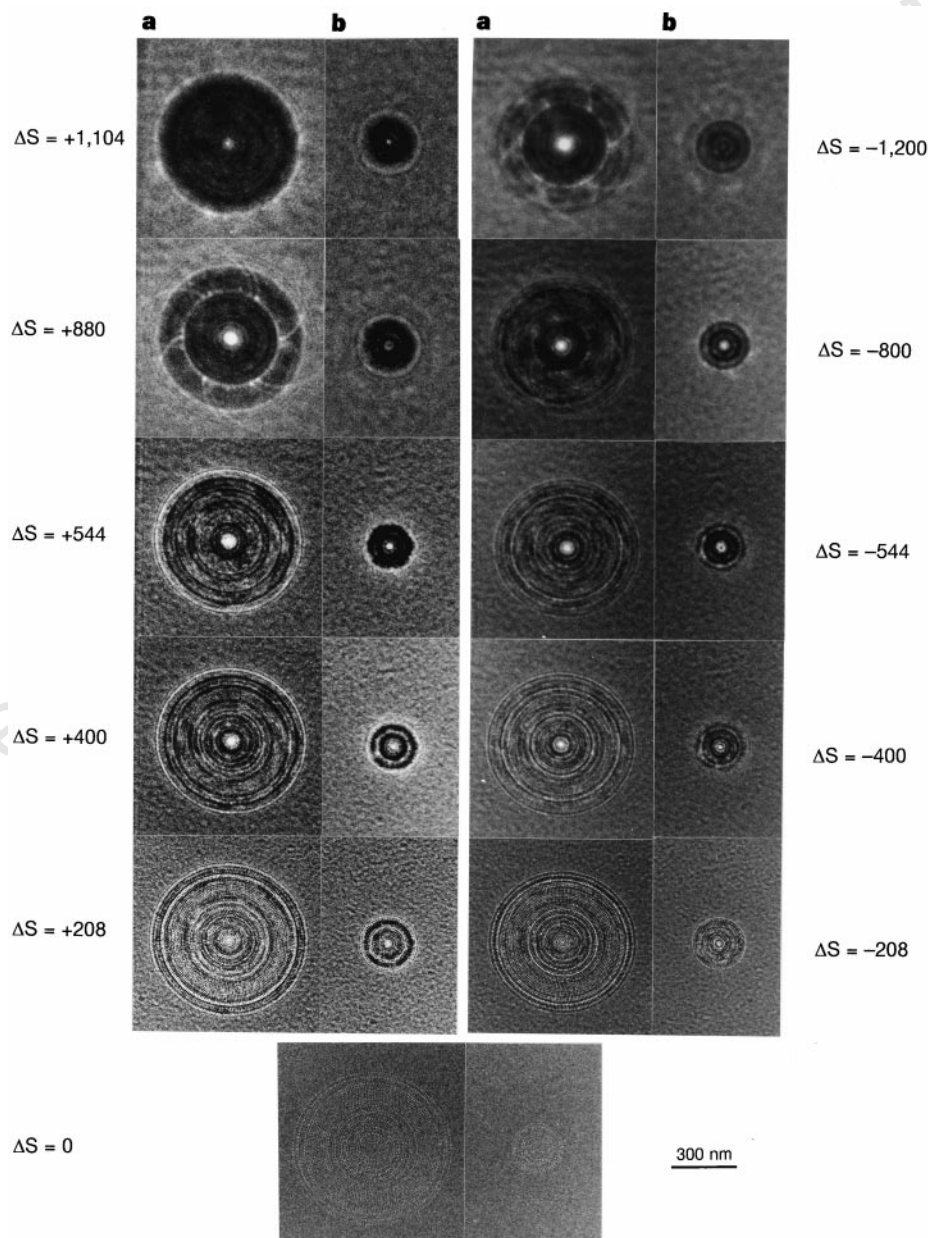


Figure 3 Electron micrographs of a through focal series of the two lenses in Fig. 2 observed in the CTEM normal imaging mode. The images at each defocus are taken on the same photographic negative therefore the illumination conditions for the two lenses in the microscope are identical. The difference of the background contrast occurred during the photographic printing process. The nominal defocus distance for each objective fine step is 640 nm (although this is dependent on the objective lens excitation). ΔS is the number of the objective fine steps. Two lenses exhibit characteristic focal lengths. **a**, Shows the brightest spot at

$\Delta S = +880$ whereas **b** shows it at $\Delta S = +544$, that is **a** has a longer focal length than **b**. At $\Delta S = 0$, the contrast is a minimum despite being intentionally enhanced in the initial printing process. The fan-like contrast of lens **a** at $\Delta S = +880$ is due to the original patterning sequence. In order to avoid obvious asymmetry in the lens by using the same angle origin for each concentric ring pattern, ring number i started with its angular origin at i radians. However, this large-scale periodicity in the pattern was unexpectedly visible.

exit surface of the film. Any multiple of this thickness also satisfies the requirement for the lens construction if inelastic scattering and large-angle elastic scattering effects are ignored. The minimum zone width limiting the resolution is determined by the minimum spacing between the holes or the hole diameter, whichever is larger. A minimum spacing of about 10 nm can be reliably achieved with a 0.5 nm diameter electron probe of a scanning transmission electron microscope (STEM). The maximum achievable size of the structure without any elaborate patching techniques⁹ is about 1 μm . Zones beyond the third should be narrower than 10 nm and therefore will contribute to the intensity but not the resolution. Figure 2 shows two lenses with different focal lengths produced according to equation (4). The larger of the two lenses has a focal length of 1 mm and the smaller one of 0.25 mm. The contrast of the constituent holes of both lenses is clearly visible. The band of holes of similar depth between radii of 200 nm and 300 nm in the large lens is due to an aliasing effect caused by the periodicity of the zones varying from one to two times the hole spacing.

A through focal series of the PFP lenses observed in a conventional transmission electron microscope (CTEM) reveals the intensity of the electron distribution along the optical axis of each lens (Fig. 3). The minimum contrast at in-focus ($\Delta S = 0$) indicates little scattering of the incident wave amplitude. As the objective lens is further overfocused (+), the centre of each PFP lens gradually becomes brighter with varying contrast patterns of the outer zones. The intensity of the centre is a maximum at around a nominal defocus of 0.56 mm ($\Delta S = +880$) and 0.35 mm (+544) for lens a and b, respectively, then decreases at even further defoci. The discrepancy between the nominal defocus values and the expected focal lengths, particularly for the large PFP lens, can be ascribed to the immersion of the PFP lens in the magnetic field of the microscope objective lens, that is, the electrons are subject to an additional focusing effect between the Fresnel lens and the focal plane of the objective lens and by illumination with slight convergence. Operating the electron microscope without exciting the objective lens provided an approximate value of the focal length

of 0.8 mm for the larger lens, which was closer to the design value. The rest of the difference may be due to the illumination beam divergence. A similar, though less striking, effect is also observed in the underfocus (–) sequence. This indicates that these lenses behave not only as converging lenses but also as diverging lenses. Lens b seems to have multiple foci, indicated by the alternating contrast at the centre of the images through part of the defocus sequence.

On possible explanation of these artefacts is given by simulations, in which the on-axis intensity of the propagated wave from the exit plane of lenses is calculated. The simulation incorporated the effect of the metal-rich cylinder (Fig. 4a, b) as a result of the hole drilling process and that of the varying hole diameter with depth. In order to observe the coexistence of convex and concave properties in one lens by the simulation, it was necessary to include both these effects. The best qualitative fit to the experimental on-axis intensity for both large and small lenses was obtained with an additional phase change, proportional to hole depth, with a maximum value of -0.2π caused by the material between holes. This adds a concave component between the holes to the convex lens due to the holes (Fig. 4c, d).

These lenses did not show any obvious degradation of performance during observation (~ 4 h in 2 sessions) under the broad illumination of electrons in the CTEM. This is not surprising because it would take more than 90 h to reach the original (S)EM exposure dose. Even this is a pessimistic estimate because the damage process by which the holes are formed is dependent on the density of the current and not simply dose-dependent⁴, and second, the region surrounding each hole is significantly more damage resistant, due to the above-mentioned aluminium enrichment, than virgin material. Furthermore, they can be stored in an evacuated desiccator for more than 6 months and reused.

Solid-state Fresnel lenses for electrons have been demonstrated with nominal focal lengths of 0.25 mm and 1 mm. Both convergent and divergent lenses can be made. Their numerical aperture is about 10^{-3} , that is, an order of magnitude less than current magnetic lenses. We have shown the capabilities of direct-writing nanolitho-

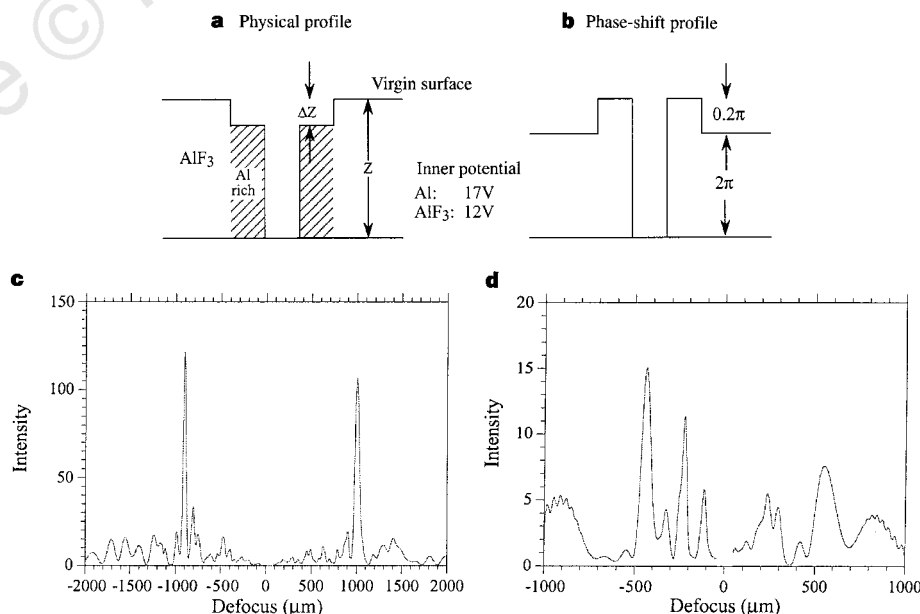


Figure 4 Schematic diagram of a cross-section of a hole made by the STEM probe. **a**, Simplified geometrical cross-section of a hole with a surrounding aluminium-rich cylinder. **b**, Corresponding net-phase profile of a plane electron wave after passing through **a**. Simulated on-axis intensity distribution of a large (700 nm) and small (250 nm) convex PFP lens as a function of defocus is shown in **c** and **d**, respectively, for 200 keV illumination. The hole diameters were varied as a function of depth up to 10 nm for the complete hole through the thickness of the

film. The influence of the proximity effect (the material between the holes after drilling) on the phase shift as modelled in **a** and **b** was included. The maximum phase shift due to this proximity effect was limited to -0.2π by setting the so-called proximity factor to 0.1. The raw intensity distribution was convolved with a gaussian distribution function to simulate blurring due to instabilities and incoherence over a defocus range of 10 μm .

graphy in inorganic resists and phase manipulation. A possible application is to X-ray optics where the required thickness of the lens could be built up by repeatedly evaporating and patterning ~ 100 nm of AlF_3 . Alignment could be retained by imaging with electrons (perhaps secondaries) emerging through the few rings of complete holes, provided stress in the film did not cause excessive distortion. This provides a potential route to new X-ray optics and the nanofabrication of photonic crystals¹⁰. More directly, an array of such lenses could be used for simultaneously patterning prescribed features of electronic devices by electron beam lithography using organic resists that are 10^4 to 10^5 times more sensitive than AlF_3 (ref. 11). There seems to be little difficulty with present equipment to make, say, 1,000 lenses on a single thin substrate. Thus, in principle, electron beam production of integrated circuits can be carried out. □

Received 2 February; accepted 16 April 1998.

1. Scherzer, O. The theoretical resolution of the electron microscope. *J. Appl. Phys.* **20**, 20–29 (1949).
2. Broers, A. N. et al. in *Electron Microscopy 1978* Vol. III (ed. Sturgess, J. M.) 343–354 (Microscopical Society of Canada, 1978).
3. Isaacson, M. & Murray, A. In-situ vaporization of very low-molecular weight resists using 1–2 nm diameter electron beams. *J. Vac. Sci. Technol.* **19**, 1117–1120 (1981).
4. Berger, S. D., Macaulay, J. M., Brown, L. M. & Allen, R. M. High resolution electron beam induced desorption. *Mater. Res. Soc. Symp. Proc.* **129**, 515–520 (1989).
5. Mochel, M. E., Humphreys, C. J., Eades, J. A., Mochel, J. M. & Petford, A. M. Electron beam writing on a 20-Å scale in metal β -alumina. *Appl. Phys. Lett.* **42**, 392–394 (1983).
6. Ito, Y., Bleloch, A. L., Paterson, J. H. & Brown, L. M. Electron diffraction from gratings fabricated by electron beam nanolithography. *Ultramicroscopy* **52**, 347–352 (1993).
7. Ito, Y., Bleloch, A. L., Granleese, S. J. R. & Brown, L. M. Electron phase gratings by electron beam nanolithography. *Inst. Phys. Conf. Ser.* **38**, 507–510 (1993).
8. Nishihara, H. & Suhara, T. *Micro Fresnel Lenses. Progress in Optics* Vol. XXIV (ed. Wolf, E.) 3–37 (Elsevier Science, Amsterdam, 1987).
9. Buckley, C. J., Browne, M. T. & Charalambous, P. Contamination lithography for the fabrication of zone plate x-ray lenses. *Proc. SPIE* **537**, 213–217 (1985).
10. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
11. Langheinrich, W. & Beneking, H. Fabrication of metallic structures in the 10 nm region using an inorganic electron beam resist. *Jpn J. Appl. Phys.* **32**, 6218–6223 (1993).

Acknowledgements. We thank S. J. R. Granleese for assistance and for the beam writing software and G. E. Rosman for his assistance and discussion on the simulation.

Correspondence and requests for materials should be addressed to A.L.B. (e-mail: ab122@hermes.cam.ac.uk).

Covalently functionalized nanotubes as nanometre-sized probes in chemistry and biology

Stanislaus S. Wong, Ernesto Joselevich, Adam T. Woolley, Chin Li Cheung & Charles M. Lieber

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

Carbon nanotubes combine a range of properties that make them well suited for use as probe tips in applications such as atomic force microscopy (AFM)^{1–3}. Their high aspect ratio, for example, opens up the possibility of probing the deep crevices⁴ that occur in microelectronic circuits, and the small effective radius of nanotube tips significantly improves the lateral resolution beyond what can be achieved using commercial silicon tips⁵. Another characteristic feature of nanotubes is their ability to buckle elastically^{4,6}, which makes them very robust while limiting the maximum force that is applied to delicate organic and biological samples. Earlier investigations into the performance of nanotubes as scanning probe microscopy tips have focused on topographical imaging, but a potentially more significant issue is the question of whether nanotubes can be modified to create probes that can sense and manipulate matter at the molecular level⁷. Here we demonstrate that nanotube tips with the capability of chemical

and biological discrimination can be created with acidic functionality and by coupling basic or hydrophobic functionalities or biomolecular probes to the carboxyl groups that are present at the open tip ends. We have used these modified nanotubes as AFM tips to titrate the acid and base groups, to image patterned samples based on molecular interactions, and to measure the binding force between single protein–ligand pairs. As carboxyl groups are readily derivatized by a variety of reactions⁸, the preparation of a wide range of functionalized nanotube tips should be possible, thus creating molecular probes with potential applications in many areas of chemistry and biology.

Among the many reactions that can be used to derivatize carboxyl ($-\text{COOH}$) functional groups, we have concentrated on nanotube chemical modification that involves the coupling of amines to the carboxyl groups at the tip ends to form amide-linked groups (Fig. 1a)⁹. The broad applicability of this coupling reaction to aqueous and non-aqueous chemistry makes it especially attractive for nanotube functionalization.

Open-ended nanotube tips are formed while shortening the tubes in an oxidizing environment before use. A transmission electron microscopy (TEM) image of a multi-walled nanotube tip end demonstrates that this process produces open ends (Fig. 1a, inset). Carboxyl groups are expected at these open ends on the basis of previous spectroscopic studies of oxidized bulk nanotube¹⁰ and graphite^{11,12} samples. Such traditional analytical techniques are nevertheless limited in their ability to verify directly the functional groups at the very end of a specific tip, because in the ideal limit there will be only a single such group ($\sim 10^{-24}$ mol).

An alternative approach for assessing the functionality at a specific nanotube tip end is to measure the adhesion force between the tip and a surface that terminates in a known chemical functionality (that is, chemical force microscopy)^{13–15}. Previous adhesion measurements carried out as a function of solution pH (force titrations) between gold-coated Si_3N_4 tips and substrates functionalized with self-assembled monolayers (SAMs) terminating in carboxyl and hydroxyl groups showed that the fraction of proton dissociation from the surface carboxyl groups could be readily monitored by the drop in adhesion force as in a classic pH titration^{14,16,17}. If carboxyl groups do indeed exist at the nanotube ends, then they also should be able to be titrated in the same way. Force titrations between pH 2 and 9 with multi-walled nanotube tips on hydroxyl-terminated SAM substrates show a well defined drop in the adhesion force at pH ~ 4.5 (Fig. 1b) that is characteristic of the deprotonation of a carboxylic acid; the mid-point of this drop (4.5) is assigned to be the pK_a . In these and all other experiments described below, the applied loads were kept below the force required for nanotube buckling^{4,5} to ensure that only the nanotube end contacted the surface. The observed decrease in adhesion force with increasing pH is also reversible for a given tip, and the transition is observed reproducibly for other tips. The absolute value of the adhesion force at low pH can vary between tips, and we believe that this reflects a variation in the number of carboxyl groups exposed at the ends of different tips. Last, the similarity of the value of the pK_a determined in our force titrations (4.5) to the bulk solution value for benzoic acid (4.2) implies that the carboxyl group is well solvated and accessible to reaction¹⁶.

To investigate the covalent modification of nanotube tips we have coupled amines (RNH_2), which yield non-ionizable or ionizable functionalities on the tips, using carbodiimide chemistry that selectively forms amide linkages only with carboxyl groups (Fig. 1a)^{9,18}. Nanotube tips modified with benzylamine—which should expose non-ionizable, hydrophobic functional groups at the tip end—yield the expected pH-independent interaction force on hydroxyl-terminated SAM substrates. This covalent modification thus eliminates the prominent pH-dependent behaviour observed with the unfunctionalized tips. Moreover, force titrations with ethylenediamine modified (that is, amine-functionalized) tips

show no adhesion at low pH and finite adhesion above pH 7. These pH-dependent interactions are consistent with our expectations for an exposed basic amine functionality that is protonated and charged at low pH and neutral at high pH. Although the observed pK_a of the nanotube-bound amine (~ 7) is reduced relative to homogeneous solution (9–10), similar behaviour has been observed in previous studies of SAM-modified Si_3N_4 tips¹⁶. Addi-

tional experiments carried out on independent tips modified using benzylamine and ethylenediamine confirm the reproducibility of these results. We believe these data thus demonstrate unambiguously that carboxyl groups are exposed at the ends of nanotube tips, and that these groups can be covalently modified to produce probes with very distinct chemical functionalities.

We have explored several areas where these tips, and our newly developed approach to covalent modification, can be applied. First, the use of functionalized nanotube probes for chemically sensitive imaging has been investigated using patterned SAM substrates (Fig. 2). We recorded intermittent-contact or tapping-mode images (Fig. 2b) in ethanol solution, using carboxyl-terminated nanotube tips on substrates patterned¹⁹ with squares that terminate in CH_3 groups and surrounded by COOH -terminated regions. The images show a difference in phase between the two sample areas, although there is no difference in height; the tip- COOH /sample- COOH regions show a phase lag relative to the tip- COOH /sample- CH_3 regions. Recent tapping-mode studies using Au-coated Si_3N_4 tips functionalized with SAMs have shown that phase-lag differences can be quantitatively related to differences in the adhesion forces, and thus can be interpreted in terms of a map of the chemical functionality²⁰. Because we expect¹³ (and indeed find) the adhesion force between the carboxyl-terminated nanotube tip and the COOH -terminated SAM to be greater than the CH_3 -terminated

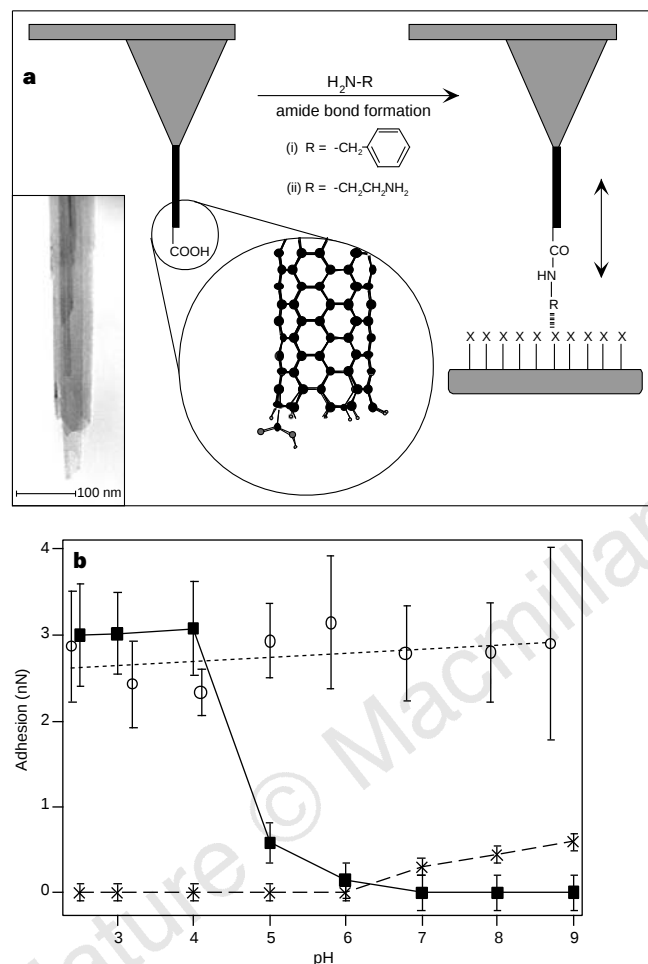


Figure 1 Preparation and characterization of functionalized carbon nanotube tips. **a**, Diagram illustrating the modification of a nanotube tip by coupling an amine (RNH_2) to a pendant carboxyl group, and the application of this probe to sense specific interactions with functional groups (X) of a substrate. The circular inset is a molecular model of a single nanotube wall with one carboxyl group at the tip end. The multi-walled nanotubes used in our studies were attached to the pyramids of gold-coated Si cantilevers ($k = 0.5\text{--}5\text{ N m}^{-1}$, Digital Instruments, Inc.) using an acrylic adhesive under the direct view of an optical microscope⁵. The as-made nanotube tips were shortened by applying a bias voltage between the tip and a sputtered Nb surface in an oxygen environment. Chemical modification of the nanotube ends was then carried out by placing a cantilever-tip assembly in a solution of 50 mM EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) (Pierce) and 5 mM of either benzylamine or ethylenediamine in 0.1 M MES (2-[N-morpholino]ethanesulphonic acid) (Sigma) buffer pH 6.0, for 2 hours. The tips were then successively washed in 0.1 M solutions of MES, NaCl (Fisher) and deionized water. The nanotubes were prepared by arc-discharge and purified by oxidation (700°C , air) until $\sim 2\%$ of the original mass remained. Inset, TEM image showing the open end of a shortened nanotube tip. **b**, Adhesion force as a function of pH between the nanotube tips and a hydroxy-terminated SAM (11-thioundecanol on gold-coated mica): filled squares, carboxyl (unmodified); open circles, phenyl (modified with benzylamine); and crosses, amine (modified with ethylenediamine). Each data point corresponds to the mean of 50–100 adhesion measurements, and the error bars represent one standard deviation.

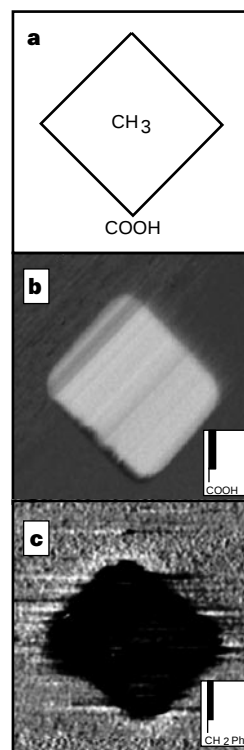


Figure 2 Chemically sensitive imaging with functionalized nanotube tips. **a**, Diagram of a patterned sample prepared by microcontact printing¹⁹ and consisting of $10\text{-}\mu\text{m}$ squares of a methyl-terminated (hexadecanethiol) SAM region surrounded by a carboxylic acid-terminated (16-mercaptohexadecanoic acid) SAM background on gold. Tapping mode phase-lag images of the patterned sample in ethanol were recorded with **b**, an unmodified nanotube tip (COOH -terminated) and **c**, a benzylamine-functionalized nanotube tip (phenyl terminated). Darker regions indicate greater phase lag; the contrast in **b** and **c** corresponds to phase variations of 2.3° and 2° , respectively. The images are $16\text{-}\mu\text{m} \times 16\text{-}\mu\text{m}$. Images and force curves were acquired with a Nanoscope III (Digital Instruments, Inc.). Imaging parameters were optimized for individual tips; typical ranges for the FESP (force modulation etched silicon probe) nanotube tips were (1) resonant frequencies, 28–33 kHz; (2) free r.m.s. oscillation amplitude, 30–90 nm, (3) set-point, 1–3 V, and (4) scan rate, 0.5–1.2 Hz.

region, these results are consistent with chemically sensitive imaging. Furthermore, when tips are covalently modified with benzylamine, which produces a hydrophobic tip that interacts more strongly with the CH₃ than with the COOH regions of the sample, the phase contrast is reversed (Fig. 2c) as expected on the basis of the change in intermolecular interactions. In addition, control experiments carried out using the same modification procedures but without the EDC coupling reagent (see Fig. 1 legend) required for covalent bond formation show the same phase contrast as the starting tips. These imaging results demonstrate that direct covalent coupling reactions on nanotube tips,

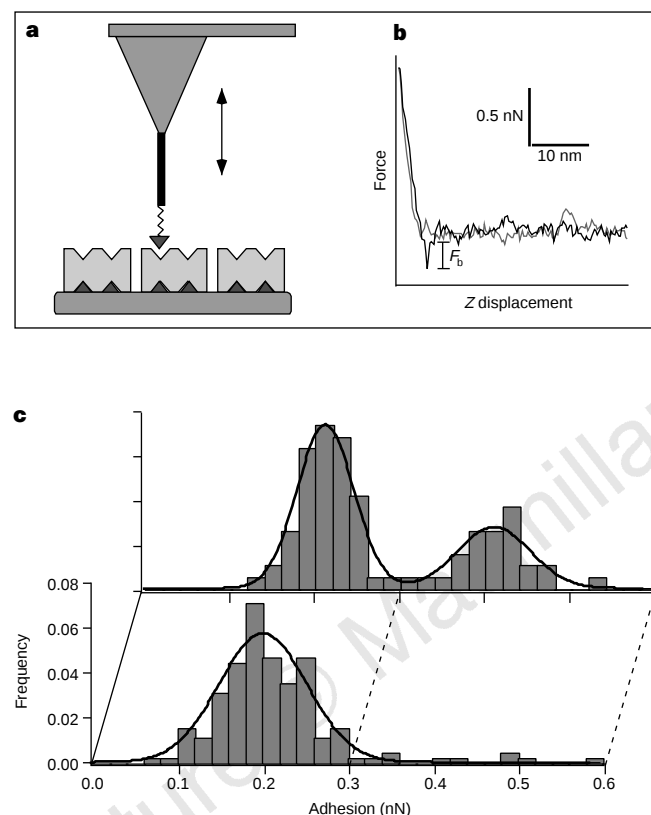


Figure 3 Ligand-derivatized nanotube tips as biological probes. **a**, Diagram illustrating a nanotube tip covalently modified with a biotin ligand (dark-grey triangle) interacting with streptavidin protein receptors (light-grey blocks); the streptavidin receptors are linked to the surface by biotin groups. The nanotube tips were modified by immersion into a solution containing 5 mM 5-(biotinamido)pentylamine (Pierce), 50 mM EDC and 0.1 M MES at pH 6.0 for 2 h. The streptavidin protein surface was formed by coating a cleaved mica substrate with 250 $\mu\text{g ml}^{-1}$ biotinamidocaproyl-labelled bovine serum albumin (Sigma) in phosphate buffer saline (PBS) pH 5.6 for 2 h, rinsing in pH 7.0 PBS, and then incubating with 30 $\mu\text{g ml}^{-1}$ streptavidin (Sigma) in pH 7.0 PBS for two hours. **b**, Representative force-displacement curve recorded with a biotin-modified nanotube tip on the streptavidin surface in pH 7.0 PBS. The binding force is indicated by F_b . **c**, Adhesion histograms representing 200–400 force-displacement curves obtained at a repetition rate of 3 Hz from two separate biotinylated nanotube tips on streptavidin-derivatized surfaces. One tip showed a single peak centred at 200 pN that corresponds to the unbinding of a single biotin-streptavidin ligand-receptor complex. The other tip exhibited a bimodal distribution peaking at 200 and 400 pN that corresponds to the unbinding of one and two biotin-streptavidin ligand-receptor complexes, respectively. The first tip showed single binding events in 36% of the measurements and no detectable binding in the remaining 64% of the curves. The second tip showed single and double binding events in 30% and 15% of the measurements, respectively. No interaction was detected within experimental error in the remaining 55% of the data. The frequency was normalized by the total number of force-displacement curves.

which we believe provide a more flexible and robust method of modification than SAMs, can be used to create chemically sensitive imaging probes. The resolution that we achieve in our experiments (Fig. 2) is limited by the technique used to create the patterned substrate. The multi-walled nanotubes used here can have diameters of 15–50 nm, but we have recently demonstrated²⁵ that lateral resolution of <3 nm can be achieved by using COOH-terminated single-walled nanotube²¹ tips on mixed monolayer/bilayer substrates.

Covalently modified nanotube tips also offer the possibility of probing biological systems at the nanometre scale. To illustrate this point we have studied the well characterized ligand-receptor interaction of biotin-streptavidin²². 5-(biotinamido)pentylamine was covalently linked to nanotube tips by the formation of an amide bond (Fig. 3a). Force-displacement measurements (Fig. 3b) made on mica surfaces containing immobilized streptavidin show well defined binding force quanta of ~ 200 pN per biotin-streptavidin pair (Fig. 3c). Control experiments carried out with an excess of free biotin (which blocks all receptor sites of the protein) in solution, and with unmodified nanotube tips showed no adhesion within the noise limits of our experiments, and thus confirm that the observed binding force results from the interaction of nanotube-linked biotin with surface streptavidin. The functionalized nanotube tips usually show only single binding events of 200 pN, although with some tips it is also possible to observe events of twice this force; we attribute such events to the simultaneous binding of two biotin-streptavidin pairs. Our measured binding force quanta agree with previous AFM studies in which biotin or avidin were attached to probe tips by the non-specific adsorption of bovine serum albumin^{23,24}. We believe that our results show that it will be possible (by using well defined covalent chemistry) to attach individual active ligands, proteins or other macromolecules in a spatially defined manner to the ends of nanotubes, and then to use these functionalized probes to create high-resolution maps of binding domains on, for example, proteins and membranes. Such experiments would be difficult using conventional tips modified either using non-specific adsorption or with SAMs.

The covalent modification of nanotube tips enables the straightforward creation of well defined probes which are sensitive to specific intermolecular interactions that define the properties of many chemical and biological systems. In addition to the directions indicated above, we believe that functionalized nanotube tips will prove especially useful for imaging self-assembled polymeric and biological materials. In particular, recent studies in which we have extended the covalent modification procedures to single-walled nanotubes²⁵ suggest the possibility of mapping functional groups with true molecular resolution. Among intriguing future applications is the use of the highly selective and robust chemistry described here to link catalysts, such as transition-metal complexes, to nanotube ends to create tools that could modify or create structures at the molecular scale. The selective functionalization of nanotube ends might also open up the possibilities of creating interconnections for electronic devices on a nanometre scale and assembling new classes of materials from nanotubes. □

Received 4 March; accepted 6 May 1998.

1. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
2. Hansma, P. K. *et al.* Tapping mode atomic force microscopy in liquids. *Appl. Phys. Lett.* **64**, 1738–1740 (1994).
3. Bustamante, C., Rivetti, C. & Keller, D. J. Scanning force microscopy under aqueous solutions. *Curr. Opin. Struct. Biol.* **7**, 709–716 (1997).
4. Dai, H., Hafner, J. H., Rinzler, A. G., Colbert, D. T. & Smalley, R. E. Nanotubes as nanoprobe in scanning probe microscopy. *Nature* **384**, 147–150 (1996).
5. Wong, S. S., Harper, J. D., Lansbury, P. T. & Lieber, C. M. Carbon nanotube tips: high-resolution probes for imaging biological systems. *J. Am. Chem. Soc.* **120**, 603–604 (1998).
6. Wong, E. W., Sheehan, P. E. & Lieber, C. M. Nanobeam mechanisms: elasticity, strength and toughness of nanorods and nanotubes. *Science* **277**, 1971–1975 (1997).
7. Keller, D. A nanotube molecular tool. *Nature* **384**, 111 (1996).
8. March, J. *Advanced Organic Chemistry* (Wiley, New York, 1992).
9. Bodanszky, M. & Bodanszky, A. *The Practice of Peptide Synthesis* 2nd edn (Springer, New York, 1994).
10. Hiura, H., Ebbesen, T. W. & Tanigaki, K. Opening and purification of carbon nanotubes in high yields. *Adv. Mater.* **7**, 275–276 (1995).

11. Evans, J. F. & Kuwant, T. Electrocatalysis of solution species using modified electrodes. *J. Electroanal. Chem.* **80**, 409–416 (1977).
12. Wandass, J. H., Gardella, J. A., Weinberg, N. L., Bolster, M. E. & Salvati, L. X-ray photoelectron and scanning Auger electron spectroscopic studies of oxidized graphite electrode surfaces. *J. Electrochem. Soc.* **134**, 2734–2739 (1987).
13. Frisbie, C. D., Rozsnyai, L. F., Noy, A., Wrighton, M. S. & Lieber, C. M. Functional group imaging by chemical force microscopy. *Science* **265**, 2071–2074 (1994).
14. Noy, A., Vezenvov, D. V. & Lieber, C. M. Functional force microscopy. *Annu. Rev. Mater. Sci.* **27**, 381–421 (1997).
15. McKendry, R., Theoclitou, M.-E., Rayment, T. & Abell, C. Chiral discrimination by chemical force microscopy. *Nature* **391**, 566–568 (1998).
16. Vezenvov, D. V., Noy, A., Rozsnyai, L. F. & Lieber, C. M. Force titrations and ionization state imaging of functional groups in aqueous solutions by chemical force microscopy. *J. Am. Chem. Soc.* **119**, 2006–2015 (1997).
17. Marti, A., Hahner, G. & Spencer, N. D. Sensitivity of frictional forces to pH on a nanometer scale: a lateral force microscopy study. *Langmuir* **11**, 4632–4635 (1995).
18. Giles, M. A., Hudson, A. Q. & Borders, C. L. Jr Stability of water-soluble carbodiimides in aqueous solution. *Anal. Biochem.* **184**, 244–248 (1990).
19. Kumar, A., Biebuyck, H. & Whitesides, G. M. Patterning self-assembled monolayers: applications in materials science. *Langmuir* **10**, 1498–1511 (1994).
20. Noy, A., Sanders, C. H., Vezenvov, D. V., Wong, S. S. & Lieber, C. M. Chemically-sensitive imaging in tapping mode by chemical force microscopy: relationship between phase lag and adhesion. *Langmuir* **14**, 1508–1511 (1998).
21. Thess, A. *et al.* Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–487 (1996).
22. Livnah, O., Bayer, E. A., Wilchek, M. & Sussman, J. L. Three-dimensional structures of avidin and the avidin-biotin complex. *Proc. Natl Acad. Sci. USA* **90**, 5076–5080 (1993).
23. Florin, E.-L., Moy, V. T. & Gaub, H. E. Adhesion forces between individual ligand–receptor pairs. *Science* **264**, 415–417 (1994).
24. Lee, G. U., Kidwell, D. A. & Colton, R. J. Sensing discrete streptavidin-biotin interactions with atomic force microscopy. *Langmuir* **10**, 354–357 (1994).
25. Wong, S. S., Woolley, A. T., Joselevich, E., Cheung, C. L. & Lieber, C. M. Covalently-functionalized single-walled carbon nanotube probe tips for chemical force microscopy. *J. Am. Chem. Soc.* (submitted).

Acknowledgements. We thank D. Vezenvov and A. Noy for discussions. C.M.L. was supported by the US Air Force Office of Scientific Research. S.S.W., E.J. and A.T.W. acknowledge fellowships from the Natural Sciences and Engineering Research Council of Canada, the Hebrew University of Jerusalem, and the Damon Runyon-Walter Winchell Foundation, respectively.

Correspondence and requests for materials should be addressed to C.M.L. (e-mail: cml@cmliris.harvard.edu).

The balance of plankton respiration and photosynthesis in the open oceans

P. J. le B. Williams

School of Ocean Sciences, University of Wales, Bangor LL59 5EY, UK

Approximately half of plant production occurs in the oceans. As oceans are open systems, a degree of imbalance between biological production and consumption can, in principle, be sustained by the import or export of organic material. Deficits in the overall oceanic budget of organic matter must be made up by import from terrestrial, freshwater and estuarine ecosystems, mainly as river-borne material. As these inputs occur at the periphery of the ocean, their contribution is largely restricted to continental-shelf waters¹. But it has been calculated²—using discrete *in vitro* observations—that in environments where net carbon fixation rates are low, respiration exceeds photosynthesis, therefore leaving the system with an organic carbon deficit. Such areas would include the central oligotrophic parts of the oceans, and it is difficult to envisage that such imbalances in these remote areas

could be sustained by organic-matter import. Here I use an analysis of depth-integrated measures of production and respiration from five open-ocean regions to show that, in the upper 100 m of the water column, biological production generally exceeds consumption. This excess is sufficient to sustain estimated organic-matter export out of these surface waters, consistent with the conclusion from simple mass-balance calculations¹ that the open oceans as a whole are not substantially out of organic carbon balance. There is no evidence of the large regional imbalances observed previously². I conclude that the form of data analysis is critical.

The biota of ecosystems are sustained by the production and respiration of organic material—the balance between these two processes being net community production (NCP). All the principal ecosystems are open and so may not necessarily be in organic (metabolic) balance. Nonetheless, it is difficult to believe that the internal processes of production and decomposition of very large ecosystems, such as the oceans, or even large regional subdivisions, are substantially out of balance—particularly in deficit. An analysis¹ of net organic balance of the oceans led to the conclusion that as a whole they were slightly (~0.6%) heterotrophic, this being mainly associated with inshore regions. The open ocean was concluded to be marginally (~0.2%) autotrophic. Such small imbalances would be beyond the resolution of field rate observations. In this respect, the recent report² that, in aquatic systems with production rates of less than 100 $\mu\text{g C l}^{-1} \text{ d}^{-1}$ (which would include large tracts of the open oceans), bacterial respiration exceeds net primary production, is unexpected³.

Overall mass-balance calculations obscure possible regional imbalances. I report here an analysis of 312 direct measurements of NCP and community respiration with wide geographical coverage (Table 1), and examine the data set in the context of the overall balance of metabolism. Except for three sets of *in situ* observations from the north central Pacific Ocean, the data are derived from *in vitro* net oxygen change and dark oxygen consumption (taken as respiration). Gross primary production (GPP) is calculated as the sum of these two processes: details of the procedures are given in refs 4–7. Combined depth profiles of NCP and GPP for six of the areas (Fig. 1) reveal no compelling evidence for the water columns under study being substantially or systematically out of balance. There is commonly an autotrophic zone of positive NCP in the upper 20 or so metres of the water column, overlying a heterotrophic zone of negative NCP, below which NCP tails off to the measurable zero rate (~0.2 $\mu\text{mol O}_2 \text{ l}^{-1} \text{ d}^{-1}$; (ref. 4 and Fig. 1). Data for the Southern Ocean stations are not shown in Fig. 1, but in all stations from this region NCP is positive throughout the observational depth range (0–80 m). It is notable that most (over 85%) of the rates of gross production in the present data set fall below 100 $\mu\text{g C l}^{-1} \text{ d}^{-1}$ (10 $\mu\text{mol O}_2 \text{ l}^{-1} \text{ d}^{-1}$, assuming a photosynthetic quotient of 1.0)—the level of net primary production (NPP) below which del Giorgio *et al.*² concluded that aquatic systems were in carbon deficit.

As the primary source of metabolic energy (photons) enters through the surface of the oceans, the balance of energy in the environment is better considered per unit area than as discrete measurements. Depth integrations, to the bottom of the observational base (Fig. 1), using a simple trapezoid procedure yielded 65 profiles (Table 1). The integrated rates fall into a general pattern.

Table 1 Depth-integrated rates of respiration, gross and net production for the five oceanic regions

Oceanic area	Number of pairs of observations	Number of integrated profiles	Number of profiles in which integrated NCP is positive	Mean depth-integrated rates ($\text{mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$)		
				GPP	NCP	NCP as % GPP
Southern Ocean	28	6	6	183 ± 80	112 ± 50	62
Northeast Atlantic Ocean	130	29	23	143 ± 66	38 ± 45	27
Arabian Sea	28	7	3	112 ± 69	–1.5 ± 38	–1.3
Mediterranean Sea	104	18	10	94 ± 76	6 ± 33	6.4
North central Pacific gyre	22	5	3	47 ± 17	–0.9 ± 43	–1.9

Productive, cold-water stations (the Southern Ocean and the North Atlantic), over the space and time frame studied, generally show a water column with high photosynthetic rates, high photosynthetic/respiration ratios (1.5–2.5) and consequent positive NCP. In these environments, NCP is a significant fraction of GPP (average 34%); that is, a substantial proportion of photosynthetic production is potentially available for export out of the space and time frame studied. The warm water areas (the Arabian and Mediterranean seas, the north central Pacific gyre) are closer to metabolic balance. The integrated rates of GPP and respiration when plotted against one another (Fig. 2a) show a pattern very different from that reported by del Giorgio *et al.*². The latter analysis gave a slope well below unity and an intercept with the 1:1 line (the point of metabolic balance, where heterotrophy matches autotrophy) well within the body of their data. In the present case, a slope (0.73) closer to unity is observed, and the point of metabolic balance ($16.7 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$) occurs at the lower end of the distribution of the observations, that is, extreme oligotrophy.

Thus, in contrast to the conclusion of del Giorgio *et al.*² that large

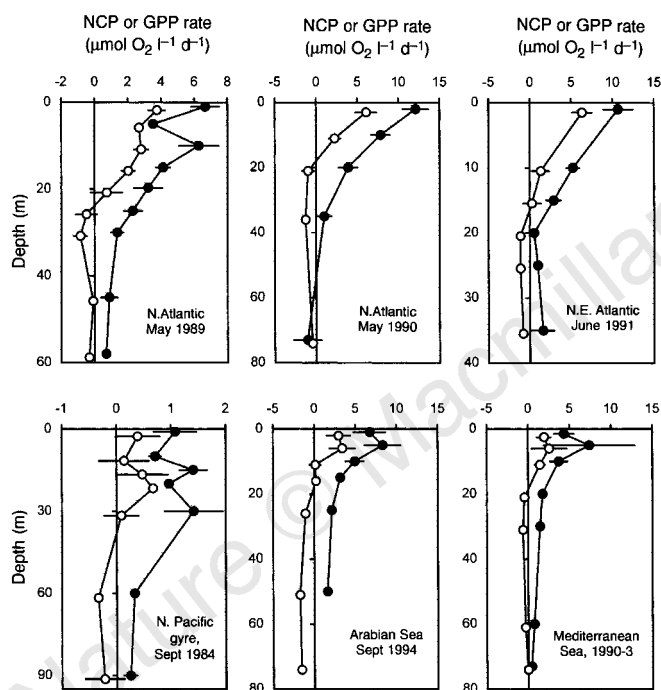


Figure 1 Composite profiles of net community production (NCP) and gross primary production (GPP) from six sets of observations. The mean rates (in $\mu\text{mol O}_2 \text{ l}^{-1} \text{ d}^{-1}$) shown are: open circles, NCP; filled circles, GPP. The bars represent the standard error of the means. The depths for the two different processes are offset slightly for clarity. The following data sets have been used in the present study. (1) The Biogeochemical Ocean Flux Study (BOFS) North Atlantic studies, as follows. (i), The BOFS North Atlantic Bloom Experiment (NABE) study; $\sim 47^\circ$ and 60°N , 20°W ; 13 May–4 June 1989. (ii) The BOFS lagrangian study; $\sim 55^\circ \text{N}$, 20°W ; 1–19 May, and 28 May–14 June 1991. (iii) The BOFS coccolithophorid study; $\sim 60^\circ \text{N}$ and 20°W in the northeast Atlantic; 16–29 June 1991. The data are all now available from the British Oceanographic Data Centre, Proudman Oceanographic Laboratory, Birkenhead L43 7RA, UK. (2) The EROS northwest Mediterranean data set. This is derived from a series of cruises in the Gulf of Lion over the years 1990–93. Details of the cruises have been published²³, and the numerical data is available from the Irish Marine Data Centre, 80 Harcourt St, Dublin 2, Ireland. (3) PRPOOS north central Pacific gyre data set; $28\text{--}29^\circ \text{N}$ and $154\text{--}155^\circ \text{W}$; 23–24 September 1984. The numerical data have been published⁴. (4) The ARABESQUE Arabian Sea data set; obtained during the post-monsoon period during September 1994¹⁸; obtained during the post-monsoon period during September 1994¹⁸; the full data set is with the British Oceanographic Data Centre. (5) The Southern Ocean data set; the data come from published papers^{5,19}. The field protocols have been published^{4–7}.

parts of the ocean are in carbon deficit, the present analysis suggests that the upper 100 m of the water column is generally autotrophic. A number of factors may have contributed to the difference in our conclusions. Different properties are compared: del Giorgio *et al.* analysed net primary production (NPP by definition is GPP minus algal respiration, obtained in their case by assuming that the ^{14}C technique measures NPP) and bacterial respiration rates; I have worked with GPP and total small community (dark) respiration. Hence there will be an offset between the two data sets, although this is unlikely to be a significant cause of the discrepancy observed. There is a potential for bias arising from the incorporation of data from freshwater and estuarine environments by del Giorgio *et al.* However, if their marine data are separated out and analysed the fit is essentially identical to the combined data set (the regressions derived from calculation of the principal axis⁸ are respectively $\log(\text{BR}) = 0.31 + 0.47\log(\text{NPP})$ and $\log(\text{BR}) = 0.34 + 0.47\log(\text{NPP})$ for the whole and the marine data, where BR and NPP are their estimates of bacterial respiration and net primary production). Over the range of rates studied, the

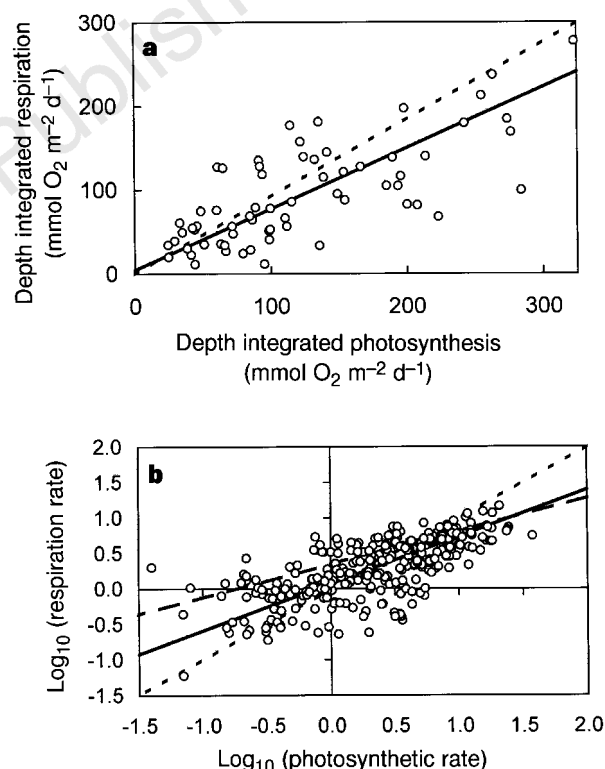


Figure 2 Comparison of rates of gross production and respiration. **a**, Regression of the depth-integrated rates of calculated gross primary production and community respiration integrated to the base of sampling. The solid fitted line ($\Sigma R = 0.73(\Sigma P) + 4.57 \text{ mmol m}^{-2} \text{ d}^{-1}$, $R^2 = 0.536$, $p = 0.02$, $n = 65$, where ΣR and ΣP are the integrated rates of respiration and photosynthesis and p is the probability of a random fit) is calculated as the first principal axis²¹. The intersection of the fitted (solid) and the 1:1 (dotted) lines is at $16.7 \text{ mmol O}_2 \text{ m}^{-2} \text{ d}^{-1}$. **b**, Log-log regression of the present discrete observations. Both rates are expressed as $\mu\text{mol O}_2 \text{ l}^{-1} \text{ d}^{-1}$. The first principal axis of the regression of the log normalized data is shown as a solid line; for comparison purposes the same analysis made on the del Giorgio *et al.*² data set is also shown (dashed line). The intersections of the 1:1 (dotted) lines with those calculated by del Giorgio *et al.* and from the present data set are respectively 15.3 and 1.63 as $\mu\text{mol O}_2 \text{ l}^{-1} \text{ d}^{-1}$ (I have used del Giorgio *et al.*'s photosynthetic quotient of 1 to interconvert the units). The equations to the fitted lines, in exponential form, are: $R = 1.18P^{0.66}$ and $R = 2.21P^{0.47}$, where R and P are the forms of respiration and production reported in the present communication and by del Giorgio *et al.*².

respiration to photosynthetic ratios in the del Giorgio *et al.* data set are 20–130% higher than the present set of observations. This may arise either from an overestimation of bacterial respiration due to the presence of algae in the filtrates, as pointed out by Geider⁹ (the upper figure is similar to the potential overestimate calculated by Geider), or because the del Giorgio data set is biased with data from more generally heterotrophic inshore sites¹. Although these factors will undoubtedly contribute to differences in the observations, the main factor giving rise to the difference in conclusion is the form of treatment of the data. del Giorgio *et al.* worked with direct property–property analysis of discrete volumetric observations, with no further treatment other than log transformation, whereas I have drawn my conclusions from depth integrated rates. A comparison of Fig. 2a with Fig. 2b (derived from the same data base) shows that the two forms of analysis give very different pictures of the distribution of the balance of plankton metabolism in the oceans. The depth-integrated approach used in the present study overcomes some of the bias in the data sets and gives a more accurate picture of the distribution of the balance of autotrophy and heterotrophy in the oceans. I have come to the conclusion that with the data sets of net primary production and respiration currently available to us, generalizations on the regional distribution of carbon balance in the oceans cannot be derived from simple regression analysis of volumetric observations.

The data in the present study come from five oceanic areas of widely varying ecotype, and so provide a basis for further generalization. They are, however, limited to the upper 100 m of the ocean, and not all seasons are covered in all of the data sets. This places a bias if one seeks to take up the challenge² to extrapolate to broader space and timescales. Temporal restriction is an obvious concern in areas showing pronounced seasonality—such as the cold water regions and the Arabian Sea. In such areas, one may expect a period of positive production followed by a period of negative net production—the latter was probably sampled in the 1990 North Atlantic lagrangian¹⁰ studies and in the Arabian Sea studies⁶. The linkage between the autotrophic and heterotrophic phases of the bloom has been examined in detail in coastal waters^{11–14}. The conclusion from these and the 1990 North Atlantic lagrangian study¹⁵ is that there is a very limited separation in time (5–10 days or less) between the two phases—with no protracted period of high heterotrophy following the autotrophic bloom. Thus, with the possible exception of the Arabian Sea data set, the lack of extensive temporal coverage would not appear to bias the data sets severely.

The important question remains of whether the estimated excess production over consumption in the upper water column is sufficient to sustain export production. Understanding of deep-water metabolism is very poor, but a number of studies have attempted to determine export production as a fraction of surface water production. The broad generalization has been made¹⁶ that 10–20% of primary production is exported from the euphotic zone, with the likelihood that a greater percentage is exported in areas of high production^{16,17}. This simple generalization would leave the cold-water areas in the present study substantially autotrophic overall (see last column of Table 1). Studies in productive areas such as the Bellingshausen Sea^{17,18} and the North Atlantic^{19–21} give a wider span of export percentages, ranging from 6 to 65% of primary production (expressed characteristically as a percentage of ¹⁴C-determined production)—the higher figures being observed during bloom periods²¹. The excess GPP over respiration in the present study for the North Atlantic averaged 27%, and 62% for the Southern Ocean. During peak periods of production (integrated rates <150 mmol O₂ m⁻² d⁻¹), the fraction of production exported below 100–150 m in the North Atlantic lay in the range 12–50%, with a mean of ~40% (refs 22, 23); out of the period of peak production, export values lie in the range 6–10% (calculated from refs 10 and 19). Thus, the observed excess production in the euphotic zone of the cold-water areas derived from the present

analysis would appear to be able to sustain the calculated export. The overall balance of production and consumption is less readily determined for the low-latitude oligotrophic areas: the upper part of the water column appears to be more finely balanced. Measured export for oligotrophic low-latitude oceanic areas (derived from ¹⁵N-determined *f*-ratios and ²³⁴Th measurements^{24–26}) and calculations from measured production and sediment-trap-derived particle fluxes^{19–22,27–29} imply export percentages in the region of 5–15%, more typically below 10%. The average of excess production (NCP) in the top 100 m as a percentage of GPP in the water columns of the Mediterranean and the north central Pacific gyre was 7%, giving no compelling grounds to conclude that there is a large metabolic imbalance in the water columns of these areas. My analysis provides no evidence to suggest that the open oceans, either as a whole or regionally, are substantially out of organic balance and thus they are not a major unrecognized sink in the global carbon budget. □

Received 16 December 1997; accepted 16 April 1998.

- Smith, S. V. & Mackenzie, F. T. C. The oceans as a net heterotrophic system: implications from the carbon biogeochemical cycle. *Glob. Biogeochem. Cycles* **1**, 187–198 (1987).
- del Giorgio, P. A., Cole, J. J. & Cimbleris, A. Respiration rates in bacteria exceed plankton production in unproductive aquatic systems. *Nature* **385**, 148–151 (1997).
- Kirchman, D. L. Microbial breathing lessons. *Nature* **385**, 121–122 (1997).
- Williams, P. J. le B. & Purdie, D. A. *In vitro* and *in situ* derived rates of gross production, net community production and respiration of oxygen in the oligotrophic subtropical gyre of the North Pacific Ocean. *Deep-Sea Res.* **138**, 891–910 (1991).
- Aristegui, J., Montero, M. F., Ballesteros, S., Basterretxea, G. & van Lenning, K. Planktonic primary production and microbial respiration measured by ¹⁴C assimilation and dissolved oxygen changes in coastal waters of the Antarctic Peninsula during the austral summer: implications for carbon flux studies. *Mar. Ecol. Prog. Ser.* **132**, 191–201 (1996).
- Robinson, C. & Williams, P. J. le B. Plankton net community production and dark respiration in the Arabian Sea during September 1994. *Deep-Sea Res.* **II** (in the press).
- Boyd, P., Robinson, C., Savidge, G. & Williams, P. J. le B. Water column and sea ice primary production during austral spring in the Bellingshausen Sea. *Deep-Sea Res.* **II** **42**, 1177–1200 (1995).
- Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, New York, 1981).
- Geider, R. J. Photosynthesis or planktonic respiration? *Nature* **388**, 132 (1997).
- Savidge, G., Boyd, P., Pomroy, A., Harbour, D. & Joint, I. Phytoplankton production and biomass estimates in the north east Atlantic ocean, May–June (1990). *Deep-Sea Res.* **II** **42**, 599–617 (1995).
- Billen, G. & Fontigny, A. Dynamics of a *Phaeocystis*-dominated spring bloom in Belgian coastal waters II Bacterioplankton Dynamics. *Mar. Ecol. Prog. Ser.* **37**, 249–257 (1987).
- Laanbroek, H. J., Verplanke, J. C., de Visscher, P. R. M. & de Vuyt, R. Distribution of phyto- and bacterioplankton growth and biomass parameters, dissolved inorganic nutrients and free amino acids during a spring bloom in the Oosterschelde basin, the Netherlands. *Mar. Ecol. Prog. Ser.* **25**, 1–11 (1985).
- van Boekel, W. H. M., Hansen, F. C., Reigman, R. & Bak, R. P. M. Lysis-induced decline of a *Phaeocystis* spring bloom and coupling with the microbial food web. *Mar. Ecol. Prog. Ser.* **81**, 268–276 (1992).
- Blight, S. P. *et al.* The phasing of autotrophic and heterotrophic plankton metabolism in a temperate coastal ecosystem. *Mar. Ecol. Prog. Ser.* **128**, 61–75 (1995).
- Williams, P. J. le B. Net production and respiration. In *The Changing Ocean Carbon Cycle: A Midterm Synthesis of the Joint Global Ocean Flux study* (Hanson, R. B., Ducklow, H. W. & Field, J. G.) (IGBP Ser. 3, Cambridge Univ. Press, in the press).
- Berger, W. H., Smetacek, V. S. & Wefer, G. In *Productivity of the Ocean; Past and Present* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 1–34 (Wiley, Chichester, 1989).
- Betzer, P. R. *et al.* Primary productivity and particle fluxes on a transect of the equator at 153° W in the Pacific Ocean. *Deep-Sea Res.* **I** **31**, 1–11 (1984).
- Shimmield, G. B., Ritchie, G. D. & Fileman, T. W. The impact of marginal ice zone processes on the distribution of ²¹⁰Pb, ²¹⁰Po and ²³⁴Th and implications for new production in the Bellingshausen Sea, Antarctica. *Deep-Sea Res.* **II** **42**, 1313–1335 (1995).
- Boyd, P. W. & Newton, P. P. Does planktonic community structure determine downward particulate organic carbon flux in different oceanic provinces? *Deep Sea Res.* **I** (in the press).
- Boyd, P. & Newton, P. Evidence of the potential influence of planktonic community structure on the interannual variability of particulate organic carbon flux. *Deep Sea Res.* **I** **42**, 619–639 (1995).
- Buesseler, K. O. The de-coupling of production and particulate export in the surface ocean. *Glob. Biogeochem. Cycles* **12**, 297–310 (1998).
- Carlson, C. A., Ducklow, H. W. & Michaels, A. F. Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso sea. *Nature* **371**, 405–408 (1994).
- Lefèvre, D. *et al.* Review of gross community production, primary production, net community production and dark community respiration in the Gulf of Lions. *Deep-Sea Res.* **II** **44**, 801–832 (1997).
- Eppley, R. W. In *Productivity of the Ocean; Past and Present* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 85–97 (Wiley, Chichester, 1989).
- Murray, J. W., Downes, J. N., Strom, S., Wei, C.-C. & Jannasch, H. W. Nutrient assimilation, export production and ²³⁴Th scavenging in the eastern equatorial Pacific. *Deep-Sea Res.* **I** **36**, 1471–1489 (1989).
- Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. VERTEX: carbon cycling in the north east Pacific. *Deep-Sea Res.* **I** **34**, 267–285 (1987).
- Jochem, F. J. & Zeitzschel, B. Production regime and phytoplankton size structure in the tropical and subtropical North Atlantic in spring 1989. *Deep-Sea Res.* **II** **40**, 494–520 (1993).
- Claustre, H. & Marty, J. C. Specific phytoplankton biomasses and their relation to primary production in the tropical North Atlantic. *Deep-Sea Res.* **I** **42**, 1475–1493 (1995).
- Michaels, A. F., Bates, N. R., Buesseler, K. O., Carlson, C. A. & Knap, A. H. Carbon cycle imbalances in the Sargasso Sea. *Nature* **372**, 537–540 (1994).

Acknowledgements. I thank C. Robinson, D. Thomas, J. Cole, H. Kennedy, D. Kirchman and E. Naylor for comments.

Correspondence and requests for materials should be addressed to the author (e-mail: oss074@sos.bangor.ac.uk).

Influence of subglacial geology on the onset of a West Antarctic ice stream from aerogeophysical observations

R. E. Bell*, D. D. Blankenship†, C. A. Finn‡, D. L. Morse†, T. A. Scambos§, J. M. Brozena|| & S. M. Hodge¶

* Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

† Institute for Geophysics, University of Texas at Austin, Austin, Texas 78759, USA

‡ US Geologic Survey, Denver, Colorado 80225, USA

§ National Snow and Ice Data Center, Boulder, Colorado 80309, USA

|| Naval Research Laboratory, Washington DC 20375, USA

¶ US Geologic Survey, Tacoma, Washington 98406, USA

Marine ice-sheet collapse can contribute to rapid sea-level rise¹. Today, the West Antarctic Ice Sheet contains an amount of ice equivalent to approximately six metres of sea-level rise, but most of the ice is in the slowly moving interior reservoir. A relatively small fraction of the ice sheet comprises several rapidly flowing ice streams which drain the ice to the sea. The evolution of this drainage system almost certainly governs the process of ice-sheet collapse^{2–5}. The thick and slow-moving interior ice reservoir is generally fixed to the underlying bedrock while the ice streams glide over lubricated beds at velocities of up to several hundred metres per year. The source of the basal lubricant—a water-saturated till^{6,7} overlain by a water system⁸—may be linked to the underlying geology. The West Antarctic Ice Sheet rests over a geologically complex region characterized by thin crust, high heat flows, active volcanism and sedimentary basins^{9–16}. Here we use aerogeophysical measurements to constrain the geological setting of the onset of an active West Antarctic ice stream. The onset coincides with a sediment-filled basin incised by a steep-sided valley. This observation supports the suggestion^{3,17} that ice-stream dynamics—and therefore the response of the West Antarctic Ice Sheet to changes in climate—are strongly modulated by the underlying geology.

We use satellite imagery to identify streaming ice in central West Antarctica¹⁸ (Fig. 1). The 20-km-wide region, with distinct margins bordering featureless ice (Fig. 2a), is dominated by flow-parallel banding and surface undulations, characteristic of streaming ice^{19,20}. Streaming has been confirmed by GPS (Global Positioning System) measurements of surface ice velocity along a seismic line crossing the southern margin (Fig. 2c)²¹. New aerogeophysical data support the ice-streaming identification. Airborne ice-penetrating radar profiles show disrupted internal layering between the identified margins (Fig. 3). In the satellite image, the flow-parallel bands begin downslope (west) of a prominent, dark 'S'-shaped feature. Airborne laser altimetry (Fig. 2b) shows that this 'S' is caused by a 100-m-high escarpment in the ice surface which marks a significant break in ice surface slope. Upstream (east) of the 'S', the ice surface has a mean slope over 25 km of 7 m km⁻¹ which decreases to 3 m km⁻¹ downstream. The corresponding driving stress for ice-sheet flow decreases from over 100 kPa upstream of the 'S' to 60 kPa downstream. We interpret the 'S' as the onset of streaming, and will refer to the portion of the ice stream downstream of this 'S' as the onset region.

Ice-penetrating radar data (Fig. 2d) shows that the onset of streaming rests over low-lying southwest-dipping subglacial topography. The onset 'S' is bounded by a plateau to the north and a prominent ridge to the south. Two kilometres west (downstream) of the 'S', the ice stream flows in a subglacial valley bounded by steep

walls (10° or greater) (Fig. 4a). This valley has a maximum depth of ~1,900 m below sea level which shallows to 1,300 m below sea level at the western edge of the survey area. Both the aeromagnetic and airborne gravity data reveal contrasting geological structures beneath the onset region. An elongate east–west-orientated 50 nT magnetic low is centred over the region (Fig. 2f) coincident with an elongate east–west-orientated free-air gravity low (Fig. 2e). The gravity low is ~25 km across, with amplitudes as large as 50 mGal below the regional value. These anomalies appear to outline the ice-stream margins identified by satellite imagery (Fig. 4b). A seismic profile (Fig. 2c) shows that the southern boundary of the ice stream, 10 km to the west of the aerogeophysical study area, is coincident with low-velocity material interpreted as a sedimentary basin²¹.

We hypothesize that geological structures control the position of both the ice-stream margins and onset; this can be tested with potential field models. To constrain the geological structures, we constructed a model of the magnetic and gravity anomalies along a profile perpendicular to the ice stream downstream of the onset 'S' (profile B, Fig. 2). The simplest model which fits the potential field data consists of three bodies interpreted as three geological units (Fig. 4c). The density (2,690 kg m⁻³) and magnetization (3.8 A m⁻¹) of the prominent basement ridge south of the ice stream are equivalent to gabbros or highly compacted basalt flows. North of 30 km along profile B, the basement (or lower geological unit) is characterized by material of lower density

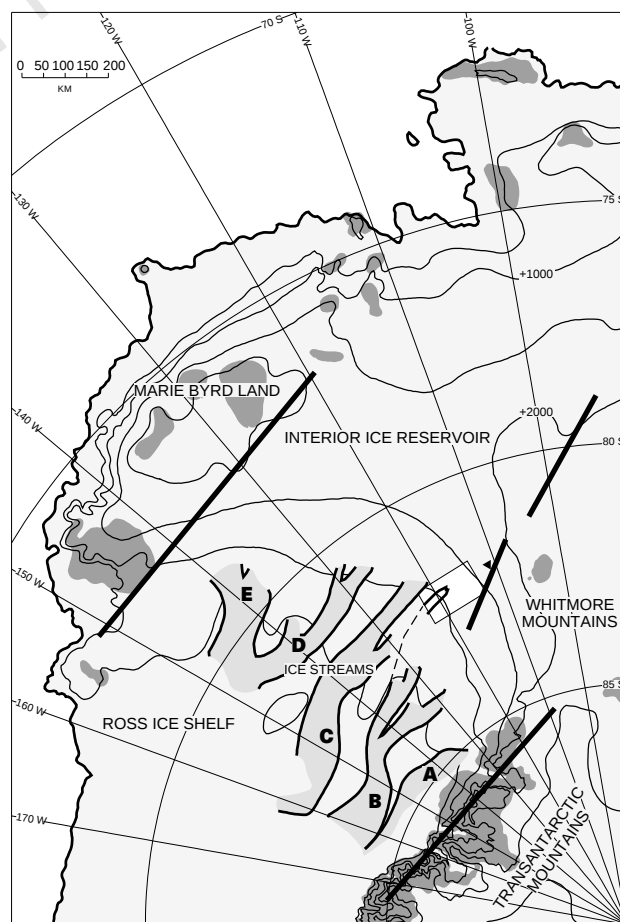


Figure 1 Map of the West Antarctic Ice Sheet. 500-m surface elevation contours are shown²³. The intermediate shading indicates the ice streams²⁴. The white box is the location of the satellite image (Fig. 2). The onset region studied here (bold outline within the white box) and its proposed flow boundaries (dashed) are shown. The onset region is connected to Ice Stream B as suggested by a recent water piracy hypothesis²⁵. Heavy lines delineate the approximate flanks of the West Antarctic rift system, and the small triangle locates an active volcano⁵.

($2,500 \text{ kg m}^{-3}$) and magnetization (1.3 A m^{-1}); this material may possibly be interlayered volcanic flows and sediments or volcanoclastic sediments. A geological model which contains only these two basement units does not adequately match the observed gravity and magnetic anomalies (r.m.s. residuals of 8.0 mGal and 15 nT for gravity and magnetics, respectively): a third geological unit is necessary to match the observed anomalies.

This third unit is constrained to correspond to sedimentary rocks by the seismic observation of $600\text{--}1,500 \text{ m}$ of low-velocity material beneath the ice stream 20 km to the west²¹. Assignment of a geologically reasonable but low density ($2,130 \text{ kg m}^{-3}$) and zero magnetization to the unit resulted in an average thickness of $1,000 \text{ m}$ of sedimentary rocks extending from 11 km to 35 km along profile B. Although the densities and magnetizations are poorly constrained, this model shows the minimum thickness of low-density material necessary to match the potential field observations (r.m.s. residuals are 2.9 mGal for the gravity and 13 nT for the

magnetics). The possible sediment thickness ranges from $1,000$ to $2,400 \text{ m}$, using the range of geologically reasonable densities ($2,130\text{--}2,400 \text{ kg m}^{-3}$). The lateral boundaries of the sedimentary basin are constrained by this modelling to within $\pm 1.5 \text{ km}$.

The modelling of profile B strongly suggests that the lateral boundaries of the ice stream near the onset of streaming are coincident with the margins of a sedimentary basin. Sedimentary basins are often long, linear features bounded by elevated topography on one or both sides. From the seismic line and the modelling along profile B we interpret this sedimentary basin as being an elongate feature orientated east–west, roughly coincident with the ice-stream margins. The unique potential field signature of this sedimentary basin is the coincident gravity and magnetic lows whose edges approximately delimit the lateral boundaries of the basin. This coincidence ceases upslope of the onset ‘S’ where the gravity and magnetic trends have significantly different orientations (Figs 2 and 4b). The question arises as to whether the sedimentary basin—which beneath the onset region contains a minimum of

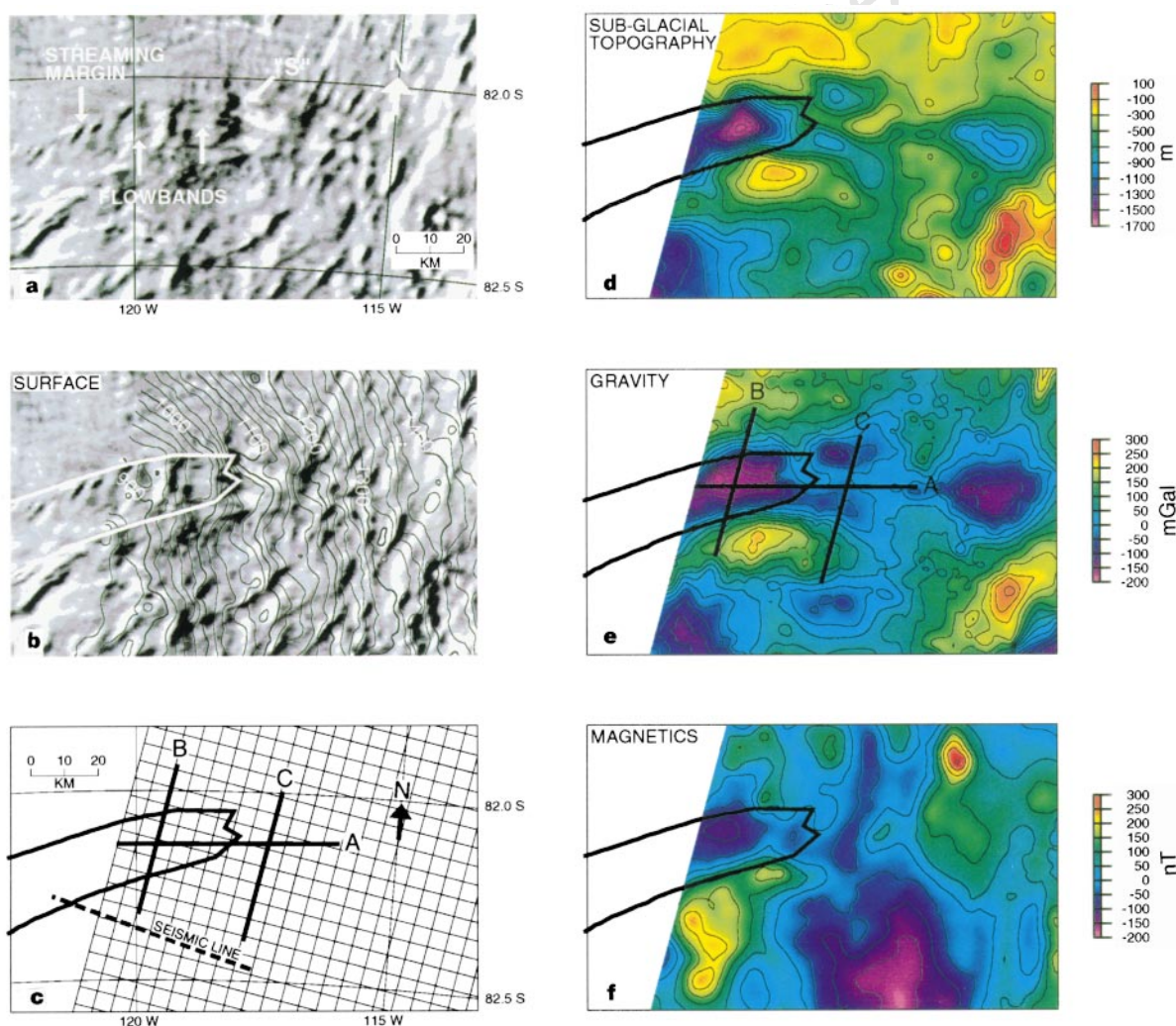


Figure 2 Aerogeophysical maps and satellite images of the onset region. See ref. 5 for a description of the acquisition of these data. **a**, AVHRR (Advanced Very High Resolution Radiometer) image. The image is $145 \times 90 \text{ km}$ and is orientated in an approximately east–west direction with illumination from the southeast. **b**, Surface elevation of the onset region from airborne laser altimetry superimposed on AVHRR image. The ice surface elevation was constructed from laser ranges to the ice surface and precise differential carrier phase GPS aircraft positions²⁶. The bold white line indicates the ice-stream margins interpreted from the AVHRR imagery. **c**, Location of aerogeophysical tracklines, profile A, profile B, profile C

and the seismic line of Anandakrishnan and others²¹. Profiles A, B and C are flightlines. The elevated ice surface velocities ($>60 \text{ m yr}^{-1}$; ref. 21) are west of the ice-stream margin interpreted from the AVHRR image. **d**, Subglacial topography map. Constructed from radar ice-thickness measurements and surface elevations from GPS-positioned laser ranges. Contour interval is 100 m . **e**, Free-air gravity map. Contour interval is 5 mGal (ref. 26). The locations of profiles A and B are shown. **f**, Magnetic anomaly map. The aeromagnetic data corrected for diurnal signals and the reference field. Contour interval is 50 nT (ref. 27).

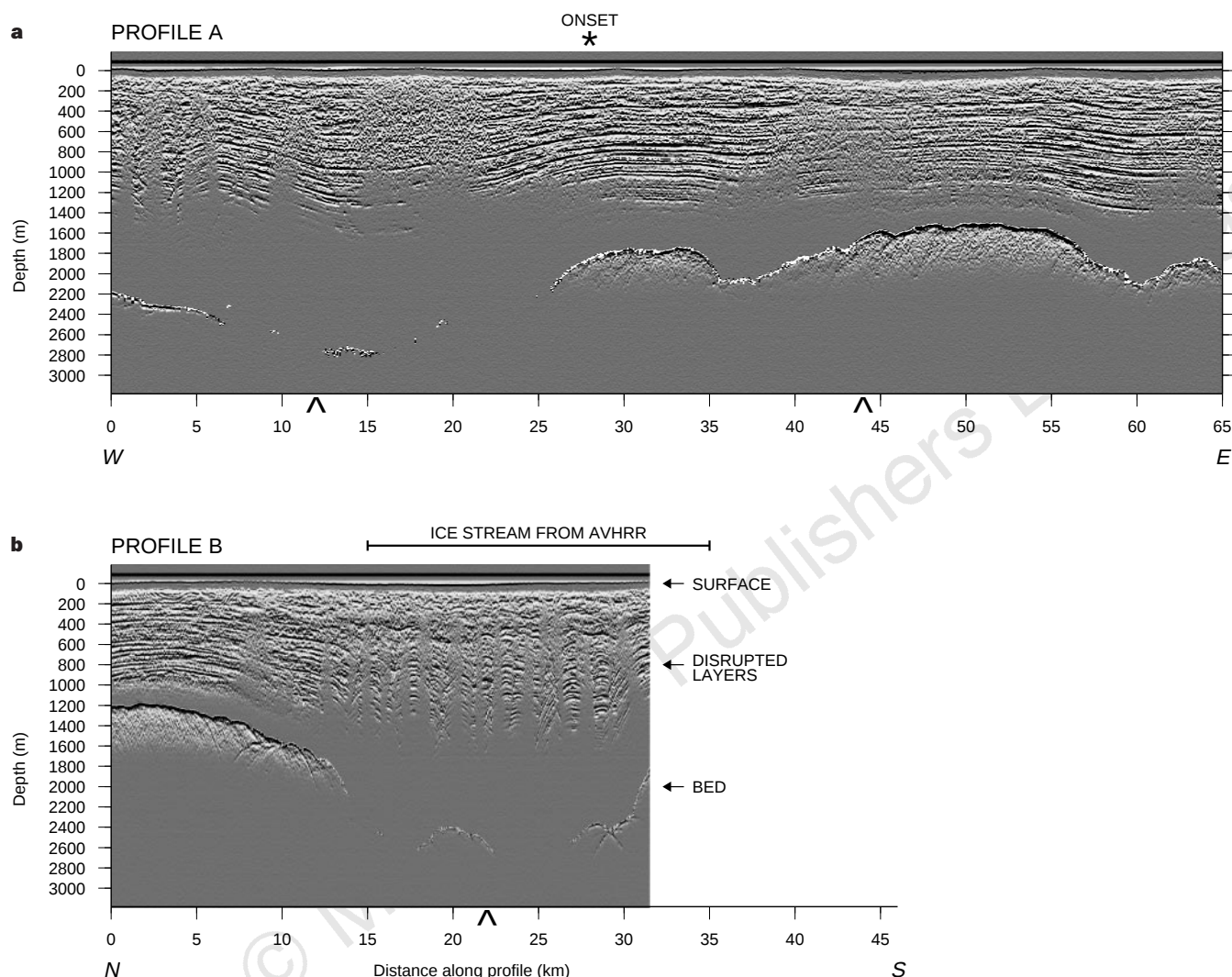


Figure 3 Radar profiles. **a**, Profile A ice-penetrating radar image along ice flow. Ice surface is the uppermost undulating dark band; 1,847 picks of bed position are marked by the lowermost sequence of white pixels. The symbol at 10.6 km indicates the intersection of profile B, and the symbol at 44 km indicates the

intersection of profile C. **b**, Profile B ice-penetrating radar image across ice flow including 793 picks of bed position. This particular radar image does not extend the full width of profile B. The symbol at 22 km indicates the intersection of profile A.

1 km of low-density low-magnetism material—terminates close to the onset 'S' as suggested by the diverging gravity and magnetic trends. Construction of a potential field model across the onset region and parallel to the ice flow (profile A in Fig. 2) would be an approach to addressing this question, but the bounding steep topography with variable density makes construction of a simple 2 (or 2.5) dimensional model impracticable. We have modelled a third profile (profile C, Fig. 4d) 10 km upstream of the onset 'S' parallel to profile B. The potential field anomalies along this profile cannot be adequately matched with a sedimentary basin. The gravity anomaly can be matched using the same two-body basement structure used in profile B but without a sedimentary basin, while the magnetic anomaly can be closely matched using a single magnetization for the entire profile, suggesting a change in the magnetic source in the region (r.m.s. residuals are 2.3 mGal for the gravity and 17.5 nT for the magnetics).

We interpret these modelling results as indicating that there is a close correlation between the margins of the ice stream and the sedimentary basin, and that the upstream edge of the sedimentary basin is closely coincident spatially with the onset of streaming. Given that the boundaries of the geological structures are generally coincident with the ice-stream boundaries, we believe that the position of this sedimentary basin controls the position of the ice-

stream margins and onset. Thin layers of sediments (10–100 m) cannot be resolved with potential field modelling and may be more widely distributed; but the thicker sequence of sediments found within a sedimentary basin have the potential to focus the streaming process by providing a longer-term source of basal lubricant.

Ice streams have been previously linked to topographic troughs¹⁴ resulting from erosion²². This new evidence linking the onset of ice streaming with an underlying sedimentary basin implies that ice streams may only develop over relatively continuous sedimentary basins. Fault-bounded sedimentary basins may be pre-existing topographic troughs but the erosion beneath an ice stream will accentuate any topographic trough, possibly producing erosional unconformities and oversteepened slopes within the scoured sedimentary basin¹⁶. Furthermore, we suggest the steep slopes of the subglacial valley (over 10°) are too steep to be the result of the initial marine deposition of the sediments within the basin, and may be indicative of active scouring of the sedimentary basin by the ice stream. The erosion of the sedimentary unit beneath the ice stream could provide the fine-grained material necessary to form a lubricating basal till.

This linkage of the onset of an Antarctic ice stream to the underlying geology raises the question of how the complex geology beneath the West Antarctic Ice Sheet may modulate long-term ice-

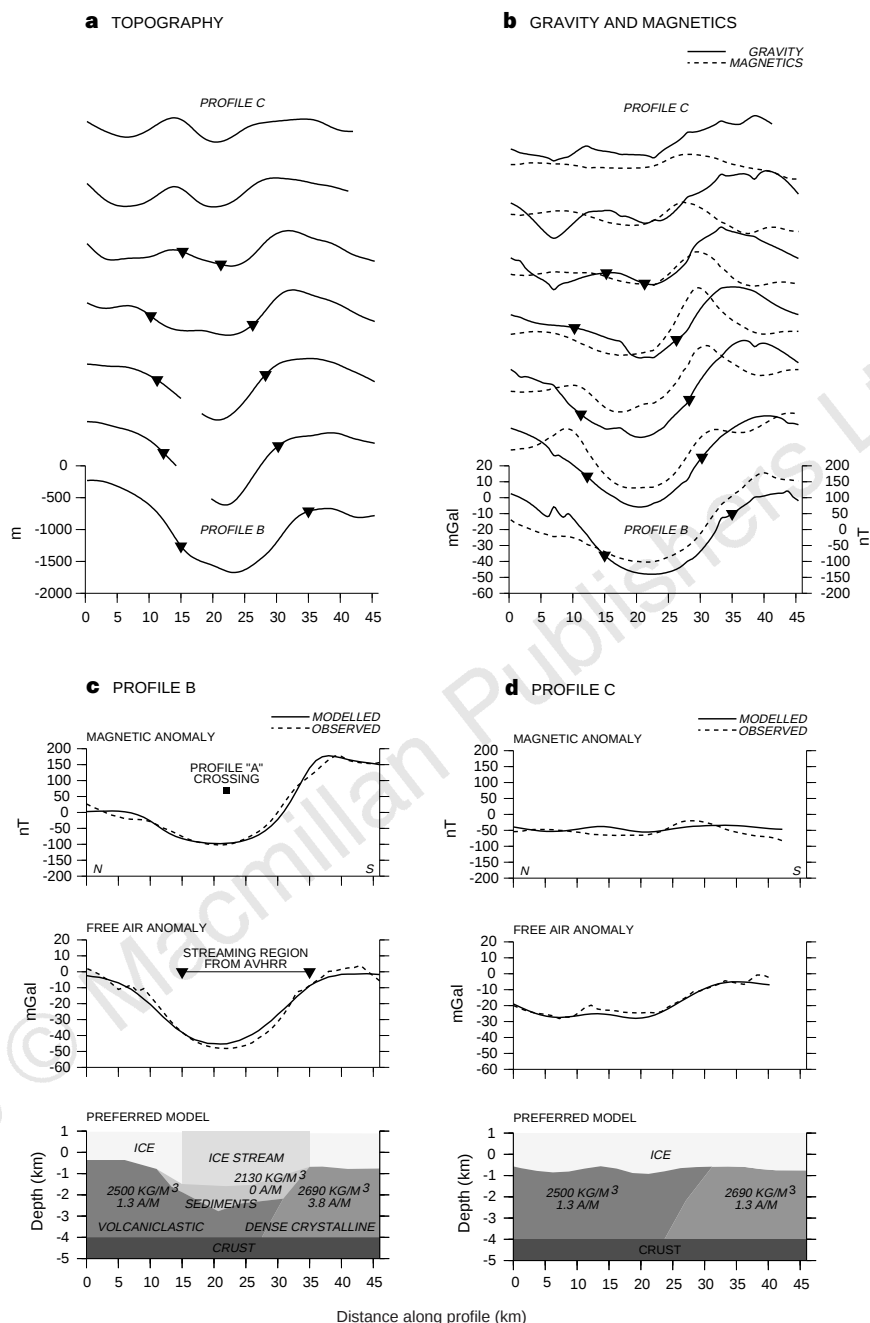


Figure 4 Potential field modelling results and stacked profiles. Ice-stream boundaries (noted with triangles) and onset were determined from the satellite imagery (Fig. 2b). **a**, Stacked subglacial topography profiles every 5.3 km beginning at profile B and extending to profile C. **b**, Stacked free-air gravity (solid lines) and magnetics (dashed lines) anomalies every 5.3 km beginning at profile B and extending to profile C. These profiles show the coincident gravity and magnetics lows which disappear just upstream of the onset. **c**, Potential field model down-

stream of onset 'S'. Shown are profile B observed gravity and magnetic values and calculated values for model with parameters below. Also shown is an illustration of the fault-bounded sedimentary basin model. Ice-stream boundaries were determined from the satellite imagery. **d**, Potential field model upstream of onset 'S'. Shown are profile C observed gravity and magnetic values and calculated values, for the model with parameters shown below.

sheet behaviour driven by global climate oscillations. We suggest that the onset of West Antarctic ice streaming can only develop over sedimentary basins (in the presence of sufficient water). The critical question is whether ice streams can only migrate along sedimentary basins or whether they can develop independently within the present interior ice reservoir. If sedimentary basins exist beneath the West Antarctic interior ice reservoir, till-lubricated ice streams should be able to rapidly drain the ice. If sedimentary basins do not exist beneath the interior ice reservoir, then a mechanism other than sediment-lubricated ice streams may be required to rapidly drain the ice sheet. We conclude that geological structures beneath the

West Antarctic Ice Sheet have the potential to dictate the evolution of the dynamic ice system, modulating the influence of changes in the global climate system. □

Received 19 March 1997; accepted 19 May 1998.

1. Fairbanks, R. G. A 17,000-year glacio-eustatic sea level record: Influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
2. MacAyeal, D. R. Irregular oscillations of the West Antarctic ice sheet. *Nature* **359**, 29–35 (1992).
3. MacAyeal, D. R. Binge/purge oscillations of the Laurentide ice sheet is a cause of the north Atlantic's Heinrich events. *Paleoceanography* **8**(6), 775–784 (1993).
4. Alley, R. B. & Whillans, I. M. Changes in the West Antarctic ice sheet. *Science* **254**, 959–963 (1991).
5. Blankenship, D. D. *et al.* Active volcanism beneath the West Antarctic Ice Sheet and implications for ice-sheet stability. *Nature* **361**, 526–529 (1993).

6. Blankenship, D. D., Bentley, C. R., Rooney, S. T. & Alley, R. B. Till beneath ice stream B.1. Properties derived from seismic travel times. *J. Geophys. Res.* **9**, 8903–8911 (1987).
7. Engelhardt, H., Humphrey, N., Kamb, B. & Fahnestock, M. Physical conditions at the base of a fast moving Antarctic ice stream. *Science* **248**, 57–59 (1990).
8. Engelhardt, H. & Kamb, B. Basal hydraulic system of a West Antarctic ice stream: constraints from borehole observations. *J. Glaciol.* **43**, 207–244 (1997).
9. Dalziel, I. W. D. & Elliot, D. G. West Antarctica: problem child of Gondwanaland. *Tectonics* **1**, 3–19 (1982).
10. Behrendt, J. C. *et al.* The West Antarctic Rift System—A Review of Geophysical Investigations 67–112 (Antarctic Res. Ser. 53, Am. Geophys. Union, Washington DC, 1991).
11. Cooper, A. K., Davey, F. J. & Hinz, K. in *Geological Evolution of Antarctica* (eds Thomson, M. R. A., Crame, J. A. & Thomson, J. W.) 285–292 (Cambridge Univ. Press, 1991).
12. Davey, F. J. *The Antarctic Continental Margin: Geology and Geophysics of the Western Ross Sea* 1–16 (Circum-Pacific Council for Energy and Resources, Houston, 1987).
13. Robertson, J. D., Bentley, C. R., Clough, J. W. & Greischar, L. L. in *Antarctic Geoscience* 1083–1090 (IUGS B4, Univ. Wisconsin Press, Madison, Wisconsin, 1982).
14. Rose, K. E. Characteristics of ice flow in Marie Byrd Land, Antarctica. *J. Glaciol.* **24**, 63–74 (1979).
15. Ten Brink, U. S., Bannister, S., Beaudoin, B. C. & Stern, T. A. Geophysical investigations of the tectonic boundary between East and West Antarctica. *Science* **261**, 45–50 (1993).
16. Rooney, S. T., Blankenship, D. D., Bentley, C. R. & Alley, R. B. Till beneath ice stream B.2. Structure and continuity. *J. Geophys. Res.* **9**(B9), 8913–8920 (1987).
17. Retzlaff, R., Lord, N. & Bentley, C. R. Airborne-radar studies: Ice streams A, B and C, West Antarctica. *J. Glaciol.* **39**(133), 495–506 (1993).
18. Hodge, S. M. & Doppelhammer, S. Onset of streaming flow of ice streams. *J. Geophys. Res.* **101**, 6669–6677 (1996).
19. Scambos, T. A. & Bindshadler, R. A. Ice flow at the confluence of two ice stream tributaries revealed by sequential satellite imagery. *Ann. Glaciol.* **17**, 177–182 (1993).
20. Bindshadler, R. A. & Vornberger, P. L. AVHRR imagery reveals Antarctic ice dynamics. *Eos* **71**, 741–742 (1990).
21. Anandakrishnan, S., Blankenship, D. D., Alley, R. B. & Stoffa, P. L. Influence of subglacial geology on the position of a West Antarctic ice stream from seismic observations. *Nature* **394**, 62–65 (1998).
22. Shabtaie, S. & Bentley, C. R. Ice-thickness map of the West Antarctic ice streams by radar sounding. *Ann. Glaciol.* **11**, 126–135 (1988).
23. Drewry, D. J. *Antarctica: Glaciological and Geophysical Folio* (Cambridge Univ. Press, 1983).
24. Shabtaie, S. & Bentley, C. R. West Antarctic ice streams draining into the Ross ice shelf: configuration and mass balance. *J. Geophys. Res.* **92**(B2), 1311–1336 (1987).
25. Anandakrishnan, S. & Alley, R. B. Stagnation of ice stream C, West Antarctica by water piracy. *Geophys. Res. Lett.* **24**, 265–268 (1997).
26. Brozena, J. M. *et al.* CASERTZ 91–92: Airborne gravity and surface topography measurements. *Antarc. J. US* **28**, 1–3 (1993).
27. Behrendt, J. C. *et al.* CASERTZ aeromagnetic data reveal late Cenozoic flood basalts in the West Antarctic rift system. *Geology* **22**, 527–530 (1994).

Acknowledgements. We thank R. Arko, M. Studinger and S. Kempf for assistance. This Letter was improved by contributions from G. Karner and C. Small. This work was supported by the US NSF.

Correspondence and requests for materials should be addressed to R.E.B. (robinb@ldeo.columbia.edu).

Influence of subglacial geology on the position of a West Antarctic ice stream from seismic observations

S. Anandakrishnan*, D. D. Blankenship†, R. B. Alley* & P. L. Stoffa†

* Earth System Science Center and Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA
† Institute for Geophysics, University of Texas–Austin, Austin, Texas 78759, USA

Ice streams drain much of the interior West Antarctic Ice Sheet and buffer the main ice reservoir from oceanic influences^{1,2}. The slow-flowing interior feeds the floating Ross Ice Shelf with ice via fast-flowing ice streams³ that are believed to modulate sea-level change through their control of inland ice storage. Understanding ice-stream behaviour, and predicting the response to climate change⁴, requires a better knowledge of the subglacial geology^{5,6}. It is known that a thawed ice-bed and high-pressure basal water are necessary, but not sufficient, conditions to cause ice streaming^{7,8}. Moreover, it has been hypothesized that a soft sedimentary bed is also required, because of its intrinsic low frictional resistance to flow⁹, and owing to its high erodibility so as to generate till that can deform and lubricate ice motion^{10,11}, or to bury rough features and smooth the bed for sliding. Here we use seismic observations to provide evidence that one margin of the upglacier part of an ice stream is directly above the boundary of a

basin with such sedimentary fill. The ice stream is within the basin and the ice outside the basin is slow-flowing. The basin fill presents an order-of-magnitude lower frictional resistance to ice flow than the subglacial material outside the basin. We conclude that the ice stream position is dependent on subglacial geology.

We identify the ice stream margin by the striking increase in flow velocity from $\sim 10 \text{ m yr}^{-1}$ to over 60 m yr^{-1} over a distance of 4.5 km (between Km 52.5 and Km 57 of Fig. 1; the Km numbers refer to distance in kilometres from the beginning of our line at Km0) that is correlative with flow-bands in satellite imagery^{12,13}. The ice flow velocities (Fig. 1b) as well as the surface elevation were determined by repeat GPS (Global Positioning System) surveying with a one-year interval between surveys. A combined reflection and refraction seismic profile along the GPS profile crossed the boundary between the slow-flowing interior ice of the West Antarctic Ice Sheet and the fast-moving ice-stream ice (Fig. 1a). The seismic data were acquired in the 1994–95 austral summer by using a towed snow-streamer with 14 Hz geophones every 25 m from 0 to 10.5 km distance from the source. The shot interval was 300 m and the line was directed east–west with higher shotpoint numbers (SP) to the west.

Common-shot gathers were collected every 300 m along the line, and the times of first arrival for the direct wave t_d and the subglacially refracted wave t_r were picked (Fig. 2 is a representative gather at location SP = Km 51.0). To interpret these data we first formed a model of the ice and subglacial structure beneath the line and then performed a least-squares inversion to determine model

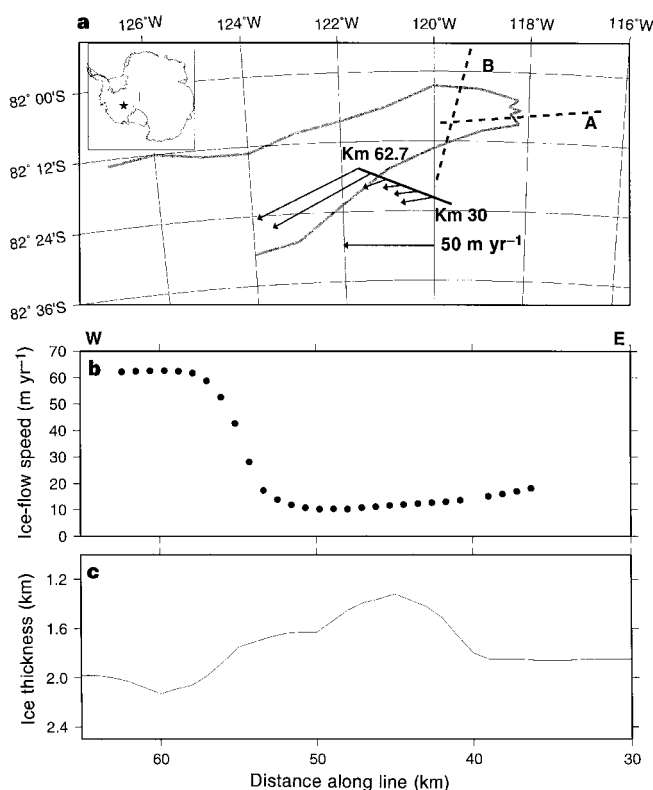


Figure 1 Location map showing the ice stream boundaries and the experiment profile line, together with ice flow speeds and ice thickness along that line. **a**, Map showing the seismic line (solid line) and selected ice flow vectors along the line (determined by repeat GPS measurements; see text). All the measured ice flow speeds are plotted in **b**. The grey curves are the boundaries and onset of ice stream C as identified by Hodge and Doppelhammer¹² (these were digitized from their Fig. 5). The two dashed lines are profiles A and B of Bell *et al.*¹³ (included here for reference). The inset is a map of Antarctica with a star marking our study area. The + symbol indicates the position of the South Pole. **b**, Ice flow speeds (m yr^{-1}); the abscissa is the position along the seismic line in kilometres; **c**, ice thickness in km.

parameters that minimized the misfit between the modelled and measured refracted-wave arrival times.

The ice-column thickness along the line was determined by creating a 'zero-offset stack section' from the ice-bottom reflection data¹⁴. The ice-bottom reflection travel-times were converted to ice thickness by using a four-layer firn seismic-velocity model and an ice-column compressional-wave (p-wave) velocity of $v_1 = 3,831 \pm 10 \text{ m s}^{-1}$, both determined by high-resolution (receiver spacing of 1 m) refraction shooting (at Km 30) as in ref. 15. Because of strain-enhanced densification, average densities in the firn layer (and the firn p-wave velocities) could differ by up to $\sim 20\%$ between the ice sheet and the ice stream (with ice-stream firn denser than ice-sheet firn at the same depths)¹⁶. This error and the uncertainty in ice velocity resulted in ice thickness determinations accurate to 15 m (Fig. 1c).

We modelled the refracted arrival travel times through our

hypothesized subglacial structures by two-dimensional ray tracing¹⁷. The arrival times are functions of shotpoint SP and offset Δ : $t_r = t_r(\text{SP}, \Delta)$. The misfit between the modelled (t_{mod}) and measured refracted arrivals, $\delta(\text{SP}, \Delta) = t_{\text{mod}} - t_r$, the r.m.s. misfit $\langle \delta \rangle$, the χ^2 value, and the goodness-of-fit q (ref. 18) were used to evaluate the models.

Initially we calculated residuals for a model of ice over a constant-velocity layer (hereafter the refracting layer is called basement). The model of ice over basement did not fit the data well (Fig. 3a). The misfit was caused by a delay in the measured t_r arrival times relative to the model (Fig. 3a). A prominent feature was a rapid increase in the absolute value of $\langle \delta \rangle$ to the west of SP = Km 56.7, marked by L, where the data were increasingly delayed relative to the model. West of SP = Km 59.7, the delay had become so large that the high-amplitude direct wave t_d arrived first and masked the refracted arrival t_r .

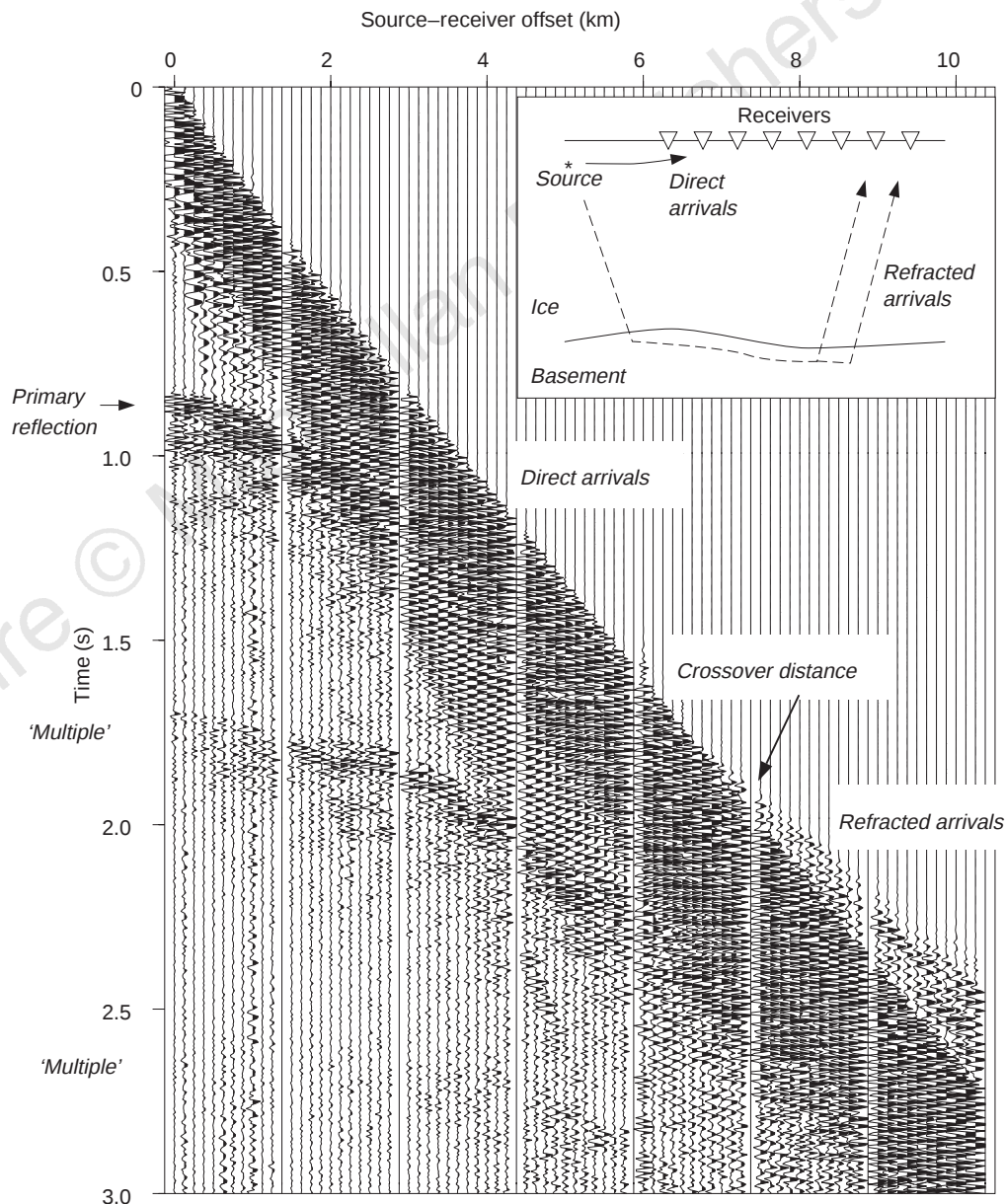


Figure 2 Common shot gather for a shot at SP = Km 51.0. Each vertical trace represents the acoustic energy arriving at that offset Δ ; the data were sampled at 1 ms resolution (for clarity, only every fifth trace is shown). The first prominent arrival at near-offsets is the 'direct arrival' (t_d) of energy that travels along the surface of the ice sheet. The break in slope of the first arrivals is at the 'crossover distance' $\Delta_c = 6.5 \text{ km}$; the first arrivals at larger offsets are from energy that has

refracted along a high-seismic-velocity basement layer beneath the ice¹⁴. Because of the $\sim 40^\circ$ incidence angle of the raypaths, there is $\sim 1.5 \text{ km}$ horizontal offset between a surface site and the corresponding illumination site at the bed. The arrival marked 'primary reflection' is the energy reflected from the bed of the ice. The inset is a diagram representing the direct and refracted energy paths.

We attempted to reduce the residuals by constructing two alternative models: one, model (1), in which refracted arrivals travel along the top of the basement, which is directly in contact with the ice, and the arrival-time delays are caused by lateral variations in seismic velocity of the basement, and the other, model (2), in which the refracted arrivals travel along the top of a fixed-velocity basement, but the delays are caused by varying thicknesses of lower-velocity sediments interposed between the ice and basement. We used a least-squares inversion¹⁹ to solve for the model parameters that resulted in the best fit between t_r and t_{mod} for each of the models. Model (1), where the inversion parameter is the basement velocity, $v_b(x)$, did not significantly improve the fit (minimum r.m.s.d. misfit $\langle \delta \rangle = 21$ ms, $\chi^2 = 6,700$, and an infinitesimal q); the v_b are well resolved, but the misfit is because of the incorrect apparent velocities of the refracted arrivals.

In model (2), the parameters are the thicknesses of the low-velocity layer $H(x)$ along the line, with the seismic velocities of both the basement and of the low-velocity layer fixed. We fixed the basement velocity at $v_b = 5.7 \text{ km s}^{-1}$, determined from a fully reversed refraction experiment between Km 42 and Km 54. This velocity is consistent with a large-offset refraction study that was conducted 30–200 km from our experiment²⁰. We introduce a layer L , of p-wave velocity v_L and thickness $H(x)$, between the ice and the basement. We place an upper bound of $v_L = 4.9 \text{ km s}^{-1}$; for

$v_L > 4.9 \text{ km s}^{-1}$, energy refracted along the top of layer L would have been observed arriving before the direct arrival, but this is not observed. A reasonable lower bound for non-dilatant sediments in central West Antarctica is $v_L = 2 \text{ km s}^{-1}$ (ref. 21). The layer p-wave velocity, v_L , was fixed at $2 \text{ km s}^{-1} \leq v_L \leq 4.9 \text{ km s}^{-1}$ and the inversion for $H(x)$ is conducted. This process was repeated for a number of v_L in this range.

The best-fit inversions yield the residuals shown in Fig. 3b and c for $v_L = 2 \text{ km s}^{-1}$ and $v_L = 4.9 \text{ km s}^{-1}$, respectively. The layer thicknesses are well resolved along the line with better than 90% confidence between Km 30 and Km 55, and better than 68% confidence between Km 55 and Km 59. An F -test for significantly different variances finds that the model with $v_L = 2.0 \text{ km s}^{-1}$ has a lower variance at the 68% significance level. Whatever the value of v_L , a thick basin must be introduced to the west of Km 51 (Fig. 4). For $x > 58 \text{ km}$, H must be greater than 400 m (the basin thickness at $x = 58 \text{ km}$) because of the absence of refracted arrivals for $SP > \text{Km } 59.7$.

Modelling of ice stream margins²² shows that when the bed friction changes abruptly from high to low friction by an order of magnitude or more, the entire margin (zone of rapid velocity change) is located on the low-friction side; for small changes in bed friction (a factor of two), the margin straddles the change and is symmetrical across the boundary. On the basis of the velocity profile

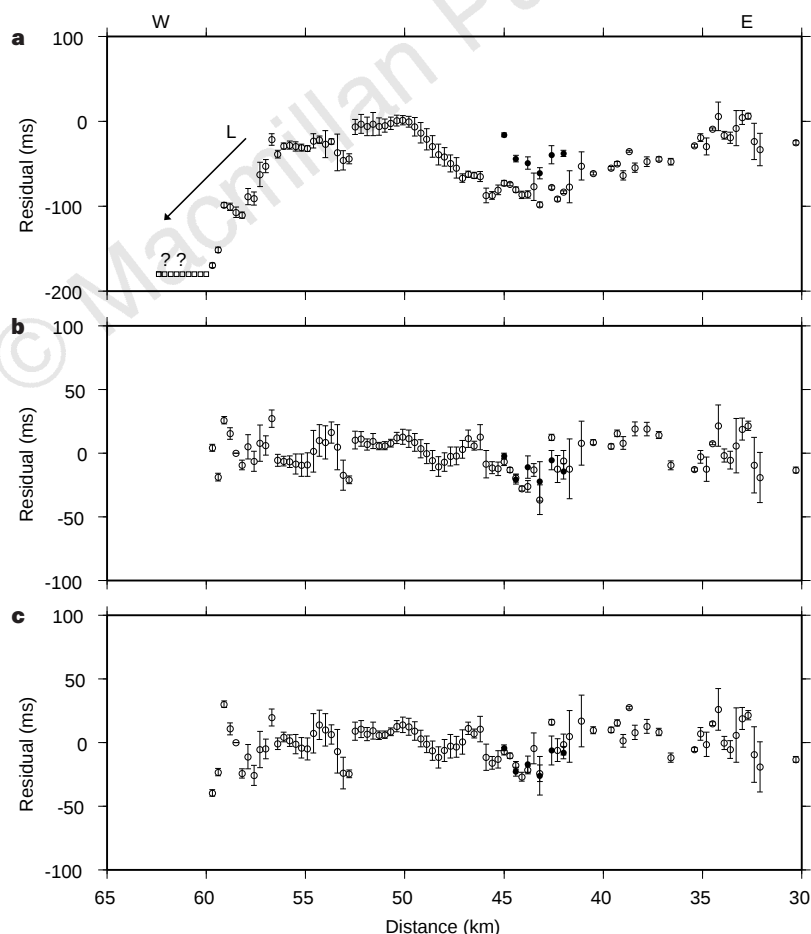


Figure 3 Mean residuals $\langle \delta(SP, \Delta) \rangle = t_{\text{mod}} - t_r$ between the model and the data are plotted at the shotpoint (SP). The error bars indicate one standard deviation. The filled circles are for 'reversed' shots where the receivers were to the west of the source; all other data were recorded with receivers to the east of the source. The open boxes (with the question marks) indicate locations where data were collected but no pre-direct-wave refracted arrivals were observed (the signal-to-noise ratios were high, so refracted arrivals would have been observed if they had been present). **a**, Residuals for a model of ice over basement as in Fig. 4, but with

no sediments. Number of data points $N = 6,195$, r.m.s. misfit $\langle \delta \rangle = 52$ ms, $\chi^2 = 1.7 \times 10^5$, and infinitesimal q . Basement velocity is $v_b = 5.7 \text{ km s}^{-1}$. The squares at $SP = \text{Km } 59.7$ and higher indicate shotpoint locations where the crossover distance was greater than the maximum offset of the recording array. **b**, As in **a**, but for ice over a sedimentary basin of $v_L = 2.0 \text{ km s}^{-1}$. r.m.s. misfit $\langle \delta \rangle = 14.9$ ms, $\chi^2 = 6,119$ and $q = 0.47$. **c**, As in **b**, but with $v_L = 4.9 \text{ km s}^{-1}$ and $\langle \delta \rangle = 15.1$ ms, $\chi^2 = 6,174$ and $q = 0.2$ (the models are shown in Fig. 4).

as well as the position of the shear margin relative to the basin boundary, we interpret the slip-friction change as at least a factor of ten²².

The ice thickness h increases by 20% across the margin into the ice stream, but the surface slope α decreases by 30% (with lower surface slopes on the ice stream) (ref. 13). Thus, the driving stress τ is lower for the fast-moving ice stream ($\tau = \rho g h \alpha$, where $\rho = 917 \text{ kg m}^{-3}$ is the density of ice, $g = 9.8 \text{ m s}^{-2}$ is the gravitational acceleration, α is the surface slope and h is the ice thickness²³). We discount a frozen bed as the cause of the slow-moving ice. The bed will be thawed if the heat flow at the bed Q (including both geothermal heat Q_{geo} and heat of sliding and heat of deformation $Q_{\text{ice}} = \tau \times u$, where u is the ice-flow speed) is high enough to maintain a bed temperature $T_{\text{B}} = -1.4^\circ\text{C}$ (the pressure-melting point for an ice-column thickness $h = 1,700 \text{ m}$, the average ice-thickness of the slow-moving ice). Using the steady-state solution due to Robin²⁴, with accumulation rate $b = 0.14 \text{ m yr}^{-1}$ and surface temperature $T_{\text{S}} = -27^\circ\text{C}$ (ref. 25), we find that for any geothermal heat flow $Q_{\text{geo}} > 35 \text{ mW m}^{-2}$, the bed will be thawed beneath our entire profile. Geothermal heat flow as low as that is unlikely in the West Antarctic; values twice as great are more likely^{5,16,26}. We justify using a steady-state calculation by noting that the regions upflow of our line are not grossly unsteady²⁷, although more work is needed. In addition, we justify *a posteriori* the conclusion that there is a thawed bed by noting that our calculation must be in error by a factor of two to negate that conclusion.

Nearly identical locations for the marked increase in ice flow speeds (at Km 52.5) and the boundary of the low-seismic-velocity subglacial basin (at Km 51) are unlikely to be a coincidence. We conclude that the low-seismic-velocity material presents a low frictional resistance to flow, and consists of easily erodible sediments similar to those that have been observed beneath ice stream B ($v_{\text{L}} = 2.0\text{--}2.3 \text{ km s}^{-1}$)²⁸; the basin has a strong negative magnetic anomaly^{13,29}, also indicative of sediments. We further suggest that these sediments are necessary to maintain rapid ice-stream flow in the upper reaches of the ice stream. The ice does not stream over the thinner, narrower low-seismic-velocity materials near Km 40 because, as demonstrated in ref. 22, narrow ice streams would be dominated by side shear, which would not permit the large velocities characteristic of streaming flow. For a potential ice stream three to four ice thicknesses wide, a decrease in basal shear stress from the driving stress τ to zero through the introduction of perfect lubrication would raise the side shear stress only to 2τ averaged over the ice thickness, producing little change in the ice

velocity in the lubricated region. Finally, a large positive gravity anomaly 10–20 km upflow of this region^{13,30} suggests that the thinner layer at Km 40 might pinch out upflow, further decreasing the likelihood of streaming.

The ice flows slowly ($10\text{--}15 \text{ m yr}^{-1}$) where sediments are intermittent. To the west of Km 51 up to the end of the line at Km 63, the low-velocity layer thickens continuously. The ice above these sediments is characteristic of ice-stream ice with higher flow-speeds ($60\text{--}70 \text{ m yr}^{-1}$), a distinct margin visible in satellite imagery, and a rapid increase in flow speed across this margin. The velocity gradient across the margin is steep, and the margin is entirely contained within the sedimentary basin, suggesting a bed with low frictional resistance within the basin. The model that best fits the data is of a 400–600-m-thick basin of $v_{\text{L}} = 2.0\text{--}2.3 \text{ km s}^{-1}$ sediments (Fig. 4). We conclude that the configuration of the ice streams, and projections about the future of the West Antarctic Ice Sheet (which is intimately tied to the ice streams) can be understood only in the context of the subglacial geology and, in particular, the distribution, thickness, and composition of sedimentary basins, all of which are poorly known. □

Received 5 March 1997; accepted 10 March 1998.

- Hughes, T. J. The West Antarctic Ice Sheet: instability, disintegration, and initiation of ice ages. *Rev. Geophys.* **13**, 502–526 (1975).
- Mercer, J. H. West Antarctic ice sheet and CO₂ greenhouse effect: a threat of disaster. *Nature* **271**, 321–325 (1978).
- Rose, K. Characteristics of ice flow in Marie Byrd Land, Antarctica. *J. Glaciol.* **24**, 63–74 (1979).
- Anandakrishnan, S. & Alley, R. B. Stagnation of ice stream C, West Antarctica by water piracy. *Geophys. Res. Lett.* **24**(3), 265–268 (1997).
- Blankenship, D. D. *et al.* Active volcanism beneath the West Antarctic ice sheet and implications for ice-sheet stability. *Nature* **361**, 526–529 (1993).
- Boulton, G. S. Theory of glacial erosion, transport and deposition as a consequence of sub-glacial sediment deformation. *J. Glaciol.* **42**(140), 43–62 (1996).
- Anandakrishnan, S. & Alley, R. B. Ice stream C, Antarctica, sticky-spots detected by micro-earthquake monitoring. *Ann. Glaciol.* **20**, 183–186 (1994).
- Alley, R. B., Anandakrishnan, S., Bentley, C. R. & Lord, N. A water-piracy hypothesis for the stagnation of ice stream C, Antarctica. *Ann. Glaciol.* **20**, 187–194 (1994).
- Boulton, G. S. & Jones, A. S. Stability of temperate ice caps and ice sheets resting on beds of deformable sediment. *J. Glaciol.* **24**(90), 29–43 (1979).
- Blankenship, D. D., Bentley, C. R., Rooney, S. T. & Alley, R. B. Seismic measurements reveal a saturated, porous layer beneath an active Antarctic ice stream. *Nature* **322**, 54–57 (1986).
- Alley, R. B., Blankenship, D. D., Bentley, C. R. & Rooney, S. T. Deformation of till beneath ice stream B, West Antarctica. *Nature* **322**, 57–59 (1986).
- Hodge, S. M. & Doppelhammer, S. K. Satellite imagery of the onset of streaming flow of ice streams C and D, West Antarctica. *J. Geophys. Res.* **101**(C3), 6669–6677 (1996).
- Bell, R. E. *et al.* Influence of subglacial geology on the onset of a West Antarctic ice stream from aerogeophysical observations. *Nature* **394**, 58–62 (1998).
- Telford, W. M., Geldart, L. P., Sheriff, R. E. & Keys, D. A. *Applied Geophysics* (Cambridge Univ. Press, 1976).
- Anandakrishnan, S., Blankenship, D. D., Alley, R. B. & Bentley, C. R. Density–depth profile determined by seismic-refraction studies: ice stream B, West Antarctica. *Ann. Glaciol.* **11**, 198 (1988).
- Alley, R. B. & Bentley, C. R. Ice-core analysis on the Siple Coast of West Antarctica. *Ann. Glaciol.* **11**, 1–7 (1988).
- Zelt, C. A. & Ellis, R. M. Practical and efficient ray tracing in two-dimensional media for rapid traveltimes and amplitude forward modeling. *Can. J. Exp. Geophys.* **24**, 16–31 (1988).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C. The Art of Scientific Computing* (Cambridge Univ. Press, 1988).
- Zelt, C. A. & Smith, R. B. Seismic traveltimes inversion for a 2-D crustal velocity structure. *Geophys. J. Int.* **108**(1), 16–34 (1992).
- Clarke, T. S., Burkholder, P. D., Smithson, S. B. & Bentley, C. R. in *The Antarctic Region: Geological Evolution and Processes* 485–493 (Terra Antarctica Publication, Siena, Italy, 1997).
- Rooney, S. T., Blankenship, D. D., Alley, R. B. & Bentley, C. R. Till beneath ice stream B. 2. Structure and continuity. *J. Geophys. Res.* **92**(B9), 8913–8920 (1987).
- Raymond, C. F. Shear margins in glaciers and ice sheets. *J. Glaciol.* **42**(140), 90–102 (1996).
- Paterson, W. S. B. *The Physics of Glaciers* 2nd edn (Pergamon, Oxford, 1981).
- Robin, G. de Q. Ice movement and temperature distribution in glaciers and ice sheets. *J. Glaciol.* **2**, 523–532 (1955).
- Whillans, I. M. & Bindshadler, R. A. Mass balance of ice stream B, West Antarctica. *Ann. Glaciol.* **11**, 187–193 (1988).
- Sclater, J. G., Jaupart, C. & Galson, D. The heat flow through oceanic and continental crust and the heat loss of the Earth. *Rev. Geophys. Space Phys.* **18**, 269–311 (1980).
- Whillans, I. M. in *The Climatic Record in Polar Ice sheets, a Study of Isotopic and Temperature Profiles in Polar Ice Sheets* (ed. Robin, G. de Q.) 70–76 (Cambridge Univ. Press, 1983).
- Rooney, S. T., Blankenship, D. D., Alley, R. B. & Bentley, C. R. in Seismic reflection profiling of a sediment-filled graben beneath ice stream B, West Antarctica. *Geological Evolution of Antarctica* (eds Thomson, M. R. A., Crame, J. A. & Thomson, J. W.) 261–265 (Cambridge Univ. Press, 1991).
- Behrendt, J. C. *et al.* CASERTZ aeromagnetic data reveal late Cenozoic flood basalts (?) in the West Antarctic rift system. *Geology* **22**, 527–530 (1994).
- Brozena, J. M. *et al.* CASERTZ 1991 and 1992: gravity and surface topography. *Antarct. J. US* **10**, 1–3 (1993).

Acknowledgements. We thank A. M. Smith, the British Antarctic Survey, S. Saustrop, UNAVCO, ASA, VXE-6, Fjord Instruments, and Landmark Graphics for assistance. We thank the US National Science Foundation for financial support.

Correspondence and requests for materials should be addressed to S.A. (e-mail: sak@essc.psu.edu).

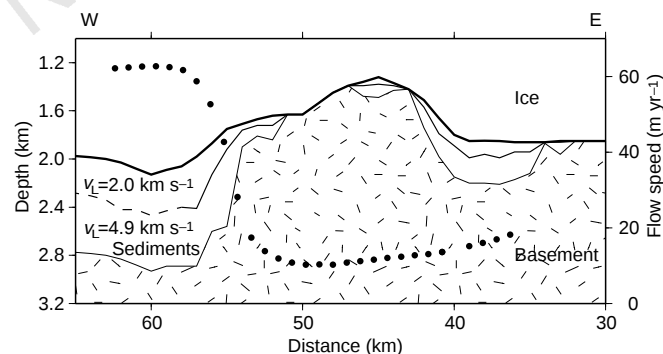


Figure 4 The final models from the ice/sediment/basement model-inversions. The ice thickness (top curve) is determined by seismic-reflection processing. The sediment layer thickness is determined by seismic-refraction processing. The layer depths for an assumed $v_{\text{L}} = 2.0 \text{ km s}^{-1}$ model (centre curve) and an assumed $v_{\text{L}} = 4.9 \text{ km s}^{-1}$ (bottom curve) are shown. To the west of Km 58 we fix the sediment thickness to the thickness value at Km 58 and indicate that with dashed lines; as discussed in the text, the sediments there are at least that thick. For reference, the ice-flow velocities along the line are superimposed, with the flow-speed axis on the right-hand side (see Fig. 1b).

A new Early Carboniferous tetrapod with a *mélange* of crown-group characters

Jennifer A. Clack

University Museum of Zoology, Downing Street, Cambridge CB2 3EJ, UK

Living (that is, crown-group) tetrapods represent the phylogenetic end-points of two lineages which diverged from each other during the mid/late Palaeozoic era. These two groups of tetrapods are the Amphibia (frogs, salamanders and caecilians), with their roots among temnospondyls^{1,2}, and the Amniota (mammals, turtles, crocodiles, birds, lizards and snakes), with their roots among anthracosaurs^{3,4}. The earliest representatives of both lineages, including a stem amniote, are known from the Viséan of East Kirkton, Scotland⁵. Here I describe a new taxon from this locality that not only combines characters of each lineage, but also represents the basal member of a third Palaeozoic group, the baphetids. The baphetids lie within the base of the crown clade of tetrapods and the morphology of the new taxon, their most primitive member, is a new benchmark for studying the polarity and evolution of crown tetrapod characters.

The specimens derive from Unit 82, the black shale member of the East Kirkton Limestone⁶. This locality has already yielded many important tetrapod fossils, including the earliest temnospondyl⁷, the two earliest anthracosaurs^{8,9}, the stem amniote *Westlothiana*¹⁰ and an aistopod¹¹. It has also yielded early terrestrial invertebrates such as eurypterids¹², scorpions¹³, myriapods¹⁴ and an opilionid¹⁵. The assemblage is considered to be one of the oldest assemblages that is indisputably terrestrial¹⁶, and is thought to have been preserved in association with a shallow lake that was subject to hydrothermal activity. At most times, the chemical-rich waters of the lake were probably inimical to vertebrate life but the shores supported a rich flora and fauna¹⁶.

Osteichthyes

Division: Crown group Tetrapoda Goodrich 1930

Order: undesignated

Family: Baphetidae Cope 1875

Eucritta melanolimnetes gen. et. sp. nov.

Etymology. Eu (Greek) true; critta (American vernacular): creature; *melano* (Greek): black; *limnetes* (Greek): living in a pool or marsh, from the black lagoon (in reference to the locality of East Kirkton).

Holotype. UMZC (University Museum of Zoology, Cambridge) T1347a and b; an almost complete articulated individual of skull length 50 mm.

Referred material. UMZC T1348 a and b, UMZC T1285 a and b and NMS (National Museums of Scotland) 1992.14 part and counterpart.

Locality. East Kirkton Quarry, near Bathgate, West Lothian, Scotland, UK.

Horizon. Unit 82 of the East Kirkton Limestone, Bathgate Hills Volcanic Formation.

Age. Brigantian, Viséan, Lower Carboniferous.

Diagnosis. Derived characters: postorbital broadly crescentic with no ventral process into orbit margin, supratemporal broadly crescentic, short snout with square or hexagonal nasals, temporal notch conspicuous with distance from its apex to the orbit margin less than the diameter of the orbit, skull table approximately square, maxillary dentition about 38–40 tooth positions including a peak at positions 7–14. Characters shared with baphetids: anteroventral embayment to orbit in (at least) larger individuals, supratemporal surrounds majority of temporal notch to the exclusion of the squamosal, narrowly triangular ventral plate to clavicle subtending

an angle of ~30° from lateral to medial. Primitive tetrapod characters: full complement of dermal roofing bones including intertemporal, supratemporal–postparietal contact, large orbits, closed palate with denticulated pterygoids meeting in the midline, broad vomerine region, fang pairs on vomers and palatines, denticulated parasphenoid with elongate triangular body, unsutured basal articulation, lozenge-shaped interclavicle, L-shaped humerus, long cleithra well ossified and expanded distally, scapulocoracoid a single ossification, cervical ribs somewhat expanded, trunk ribs scarcely curved, pentadactyl pes, ilium with both dorsal blade and posterior process, ventral armour of narrow gastralia.

The taxon is represented by four specimens, the skulls of which vary in length from 30 mm to about 90 mm. Two of these skulls preserve postcranial material. The skull roof of this species is superficially similar to that of the contemporary temnospondyl *Balanerpeton*⁷, sharing the short snout, large orbits, conspicuous temporal notch and primitive pattern of skull table bones, which show supratemporal–postparietal contact. Subtle differences distinguish them, such as the shape of the parietal plate, the position of pineal foramen, the shape of the postorbital and the extent of the contribution of the supratemporal to the margin of the temporal notch. The palate, however, does not possess the large interpterygoid vacuities that are characteristic of temnospondyls, but has the primitive closed structure that is found in other early tetrapods, including anthracosaurs and baphetids (Fig. 1). The largest specimen shows that the orbit was slightly excavated into an embayment at the anteroventral corner. A highly exaggerated version of this embayment is one of the characters that unites the other baphetids¹⁷, so *Eucritta* shows what appears to be a more primitive expression of this unusual feature.

The postcranial skeleton includes a lozenge-shaped interclavicle and an ilium bearing both a dorsal blade and a posterior process (Fig. 2), features which are also associated with other early groups including the Devonian genus *Acanthostega*¹⁸, anthracosaurs and lepospondyls¹⁹. Until recently, the postcranial skeleton morphology of baphetids (formerly known as loxommatids²⁰) was unknown, but it is now known that *Baphetes* too possesses a biramous ilium²⁰. There are also similarities in the humerus such as the shape of the entepicondyle, which is found as a quarter circle rather than as the rectangle that is found in many other tetrapods. The elongate gastralia of *Eucritta* more closely resemble those of *Baphetes* than those of *Balanerpeton*. The poorly ossified axial skeleton of *Eucritta* may likewise be characteristic of baphetids as a group, explaining why axial material in particular and postcranial material in general has so rarely been found.

The phylogeny derived from the current data set is not particularly robust (Fig. 3). This instability undoubtedly results from the extreme degree of character conflict that *Eucritta* presents, and the computed phylogeny is to that extent in agreement with initial subjective impressions. One of the more strongly supported nodes is that which places *Eucritta* as a basal baphetid (node 10). A further strong signal is that baphetids as a whole lie within the crown group (node 4) if temnospondyls and anthracosaurs are both taken to be early members of it, as is accepted by the majority of workers^{1–4} (but see ref. 21 for a conflicting view). Baphetids do not appear as a stem tetrapod plesion in any trees suggested by this analysis, but cluster more consistently with anthracosaurs than with temnospondyls.

The morphology and phylogenetic position of *Eucritta* have wide implications for studies of early tetrapods. The anterior orbital extensions of baphetids appear as small embayments in the largest skull of *Eucritta*, indicating that they may have developed by peramorphosis among more derived forms. Baphetids, like temnospondyls^{7,22} and anthracosaurs^{8,9} originated as small to medium-sized terrestrially adapted forms and each lineage convergently radiated into large, aquatic, coal-swamp piscivores. Many terrestrial forms had thus evolved by the late Viséan. The large temporal notch of *Eucritta* is difficult to interpret functionally as the

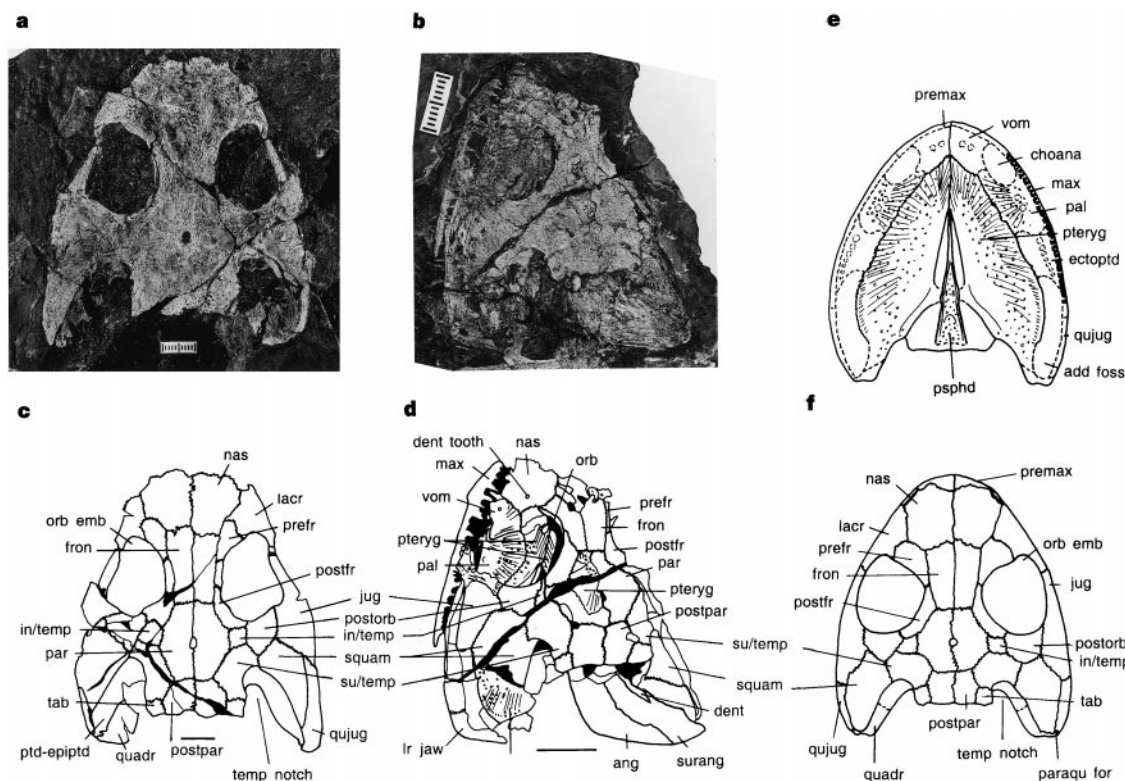


Figure 1 Skull of *Eucritta melanolimnetes*. **a**, Photograph of specimen NMS 1992.14; internal view of skull roof. **b**, Photograph of UMZC T1347b; close-up of skull in dorsal view. **c**, Composite specimen drawing from NMS 1992.14; part and counterpart (reversed from internal view). **d**, Interpretive drawing of UMZC T1347b, showing parts of palate. **e, f**, Reconstruction of palate (top) and skull roof

(bottom) of *Eucritta*; outline is based on UMZC T1347, palate on UMZC T1347 and T1348, orbital embayments from NMS 1992.14, parasphenoid from UMZC T1285. Areas coloured black represent cracks. Scale bars represent 10 mm. For abbreviations, see Methods.

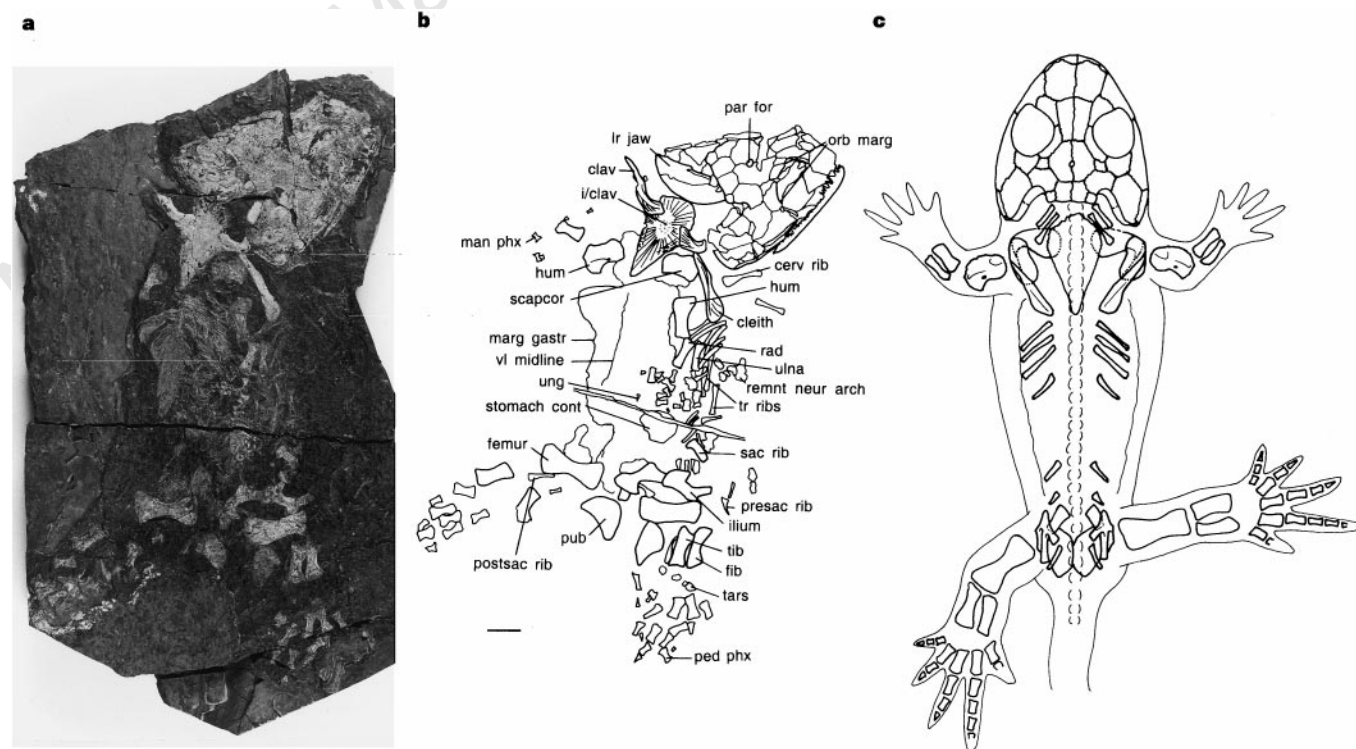


Figure 2 Body of *Eucritta melanolimnetes*. **a**, Photograph of specimen UMZC T1347a, in ventral view. **b**, Interpretive drawings of UMZC T1347a; specimen shown in ventral view. **c**, Reconstruction of postcranial skeleton based on specimen UMZC T1347. Elements from the right and left sides are freely duplicated. The carpus, tarsus and vertebrae are barely ossified; the number of

presacral vertebrae is proposed to be 24. The number of digits on the manus is unknown. *Eucritta* is a short-bodied form, with hind limbs that are conspicuously larger than the forelimbs. The dashed lines show reconstructed margins; the dotted lines show outlines of parts not visible in dorsal view. Scale bars represent 10 mm. For abbreviations, see Methods.

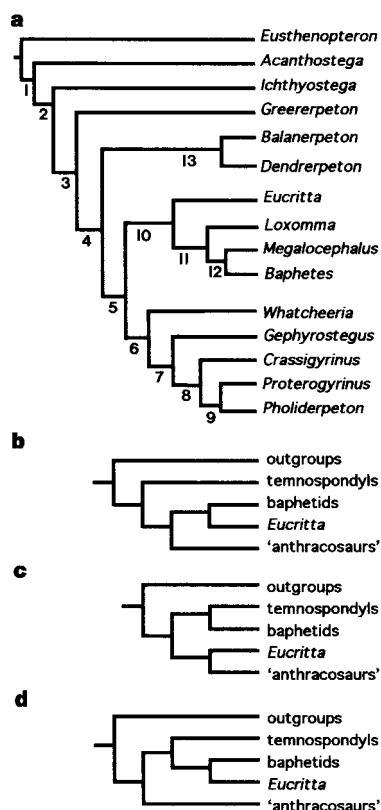


Figure 3 Phylogenetic analysis. **a**, The single most parsimonious tree. **b**, Trees placing *Eucritta* as a basal baphetid, with this clade as the sister group to anthracosaurs plus *Whatcheeria* and *Crassigyrinus* ('anthracosaurs'). **c**, Trees placing *Eucritta* with the anthracosaurs plus *Crassigyrinus* and *Whatcheeria*, and that clade as the sister group to baphetids plus temnospondyls. **d**, Tree placing *Eucritta* with the baphetids as the sister pair to temnospondyls and this clade as the pair to anthracosaurs plus *Whatcheeria* and *Crassigyrinus*. Characters supporting nodes 5, 6 and 10 are: 27, postparietal occipital exposure present; 37, suture of squamosal with supratemporal ventrally placed in embayment; 50, parasphenoid body with single median depression; 61, maxillary tooth number fewer than 30; 67, parasphenoid shagreen field situated both anterior and posterior to basal articulation; 70, premaxillary tooth number fewer than 10; 86, skull table/cheek junction square/abrupt profile; 103, trunk ribs ventrally curved; 104, length of trunk ribs more than twice the height of the neural arch plus centrum; 109, scutes tapered and elongate. (Characters 67 and 109 support their respective nodes only in their plesiomorphic state.)

stapes is unknown, but later baphetids had stapes of primitive form that are like those of the earliest tetrapods²³ and which are not associated with tympanic ears^{24–26}. The large notch in *Eucritta*, even if not part of a tympanic ear, represents a feature from which the notch in temnospondyls may have developed. *Eucritta* has a mixture of characters from three apparently diverse groups, the baphetids, temnospondyls and anthracosaurs, indicating that these three groups, which include the two crown-group lineages, diverged only a little earlier in the Carboniferous, rather than having been separate since the Devonian^{18,27}. Finally, *Eucritta* provides an important new outgroup taxon against which to examine the character polarity, and thus the phylogeny, of early crown-group members. □

Methods

Phylogenetic analysis. I analysed 15 taxa and 111 characters (87 cranial and 24 postcranial; see Supplementary information), using the phylogenetic packages PAUP²⁸ and MacClade²⁹. The taxa were chosen to represent a range of early tetrapods and members of established groups such as temnospondyls, baphetids and anthracosaurs. None of the 'lepospondyl' groups was included.

An heuristic search gave a single most parsimonious tree of length 269, consistency index 0.53 and retention index 0.57; in this tree *Eucritta* was placed as the most basal member of a clade containing three other baphetid species (*Megaloccephalus pachycephalus*, *Baphetes kirkbyi* and *Loxomma acutirhinus*). This clade emerged as the sister group to one containing the anthracosaurs species *Gephyrostegus bohemicus*, *Pholiderpeton scutigerum* and *Proterogyrinus scheeli*, as well as two other early tetrapods, the monospecific genera *Whatcheeria* and *Crassigyrinus*. That grouping proved to be the sister group to the group consisting of temnospondyls *Dendrerpeton acadianum* and *Balanerpeton woodi*. At the base of the tree, three monospecific tetrapod taxa, *Greererpeton*, *Ichthyostega* and *Acanthostega*, appeared as successively more primitive plesions. The tristichopterid fish *Eusthenopteron foordi* was used as the outgroup. A branch and bound analysis produced two equally parsimonious trees of length 269. One of these had an identical topology to that produced by the heuristic search, whereas the second was of the pattern shown in Fig. 3c.

An heuristic search made for all trees of length 270 or less resulted in recovery of a further nine trees. Of all these ten trees, three variants can be identified. Four trees (including the most parsimonious) placed *Eucritta* as a basal baphetid, with this clade as the sister group to the group consisting of the anthracosaurs plus *Whatcheeria* and *Crassigyrinus*. Five trees placed *Eucritta* with the anthracosaurs plus *Crassigyrinus* and *Whatcheeria*, and placed that clade as the sister group to baphetids plus temnospondyls. One tree placed *Eucritta* with the baphetids as the sister pair to temnospondyls, and this clade as the pair to anthracosaurs plus *Whatcheeria* and *Crassigyrinus*. Apart from the position of *Eucritta*, the major clades usually remained intact, the arrangements of genera within each clade being the main source of inconsistency. An analysis of skull characters produced similar groupings but placed *Eucritta*, the baphetids and the temnospondyls together more frequently.

A strict consensus of the ten trees places *Eucritta*, *Whatcheeria*, the temnospondyls, the baphetids and *Crassigyrinus* plus the anthracosaurs in an unresolved polytomy. A 50% majority rule consensus resolves the position of *Whatcheeria* and places it at the base of the anthracosaur/*Crassigyrinus* clade, but makes no difference to the other nodes. Computing a consensus tree is a less informative exercise than analysing the ten trees separately, and disguises the occurrence of three distinct patterns.

In the single most parsimonious tree from the heuristic search, basal nodes 2–4 each have many (more than nine) unambiguous characters supporting them, and these remain stable in the strict consensus tree. Node 5 (*Eucritta* and the baphetids plus the 'anthracosaurs') has only two unambiguous characters supporting it (characters 50 and 103), with a further five equivocal ones. This is the least strongly supported node in the cladogram, and is the one which breaks down most frequently in trees of one step longer. Five unambiguous characters (characters 27, 37, 61, 67 and 109) and a further four equivocal ones support node 10 (*Eucritta* plus remaining baphetids), whereas three unambiguous characters (characters 70, 86 and 104) and a further four equivocal ones support node 6 (the 'anthracosaurs').

Abbreviations used in figures. add foss, adductor fossa; ang, angular; cerv rib, cervical rib; clav, clavicle; cleith, cleithrum; dent, dentary; ectoptd, ectopterygoid; fib, fibula; fron, frontal; hum, humerus; i/clav, interclavicle; in/temp, intertemporal; jug, jugal; lacr, lacrimal; lr jaw, lower jaw; man phx, manual phalanx; marg gastr, margin of gastralia; max, maxilla; nas, nasal; orb, orbit; orb emb, orbital embayment; orb marg, orbit margin; pal, palatine; par, parietal; par for, parietal foramen; paraqu for, paraquadrate foramen; ped phx, pedal phalanx; postfr, postfrontal; postpar, postparietal; postorb, postorbital; postsac rib, postsacral rib; prefri, prefrontal; premax, premaxilla; presac rib, presacral rib; psphd, parasphenoid; pteryg, pterygoid; ptd-epitd, pterygoid-epiterygoid complex; pub, pubis; quadr, quadrate; qujug, quadratojugal; rad, radius; remnt neur arch, remnant of neural arch; sac rib, sacral rib; scapcor, scapulocoracoid; stomach cont, ball of stomach contents; surang, surangular; su/temp, supratemporal; squam, squamosal; tab, tabular; tars, poorly ossified tarsus; tib, tibia; tr ribs, trunk ribs; temp notch, temporal notch; ung, ungual; vl midline, ventral midline; vom, vomer.

Received 17 December 1997; accepted 2 April 1998.

1. Milner, A. R. in *The Phylogeny and Classification of the Tetrapods* (ed. Benton, M. J.) 59–102 (Clarendon, Oxford, 1988).
2. Trueb, L. & Cloutier, R. in *Origins of the Higher Groups of Tetrapods* (eds Schultze, H. & Trueb, L.) 223–314

- (Cornell Univ. Press, Ithaca, New York, 1991).
- Panchen, A. L. & Smithson, T. R. in *The Phylogeny and Classification of the Tetrapods* (ed. Benton, M. J.) 1–32 (Clarendon, Oxford, 1988).
 - Gauthier, J. A., Kluge, A. G. & Rowe, T. in *The Phylogeny and Classification of the Tetrapods* (ed. Benton, M. J.) 103–155 (Clarendon, Oxford, 1988).
 - Rolfé, W. D. I., Clarkson, E. N. K. & Panchen, A. L. (eds) Volcanism and Early Terrestrial Biotas. *Trans. R. Soc. Edinb. Earth Sci.* **84**, parts 3 and 4 (1993).
 - Rolfé, W. D. I. et al. The East Kirkton Limestone, Viséan, of West Lothian, Scotland: introduction and stratigraphy. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 177–188 (1993).
 - Milner, A. R. & Sequeira, S. E. K. The temnospondyl amphibians from the Viséan of East Kirkton, West Lothian, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 331–362 (1993).
 - Smithson, T. R. *Eldeceon rolféi*, a new reptiliomorph from the Viséan of East Kirkton, West Lothian, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 377–382 (1993).
 - Clack, J. A. *Silvanerpeton miripedes*, a new anthracosauroid from the Viséan of East Kirkton, West Lothian, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 369–376 (1993).
 - Smithson, T. R., Carroll, R. L., Panchen, A. L. & Andrews, S. M. *Westlothiana lizziae* from the Viséan of East Kirkton, West Lothian, Scotland and the amniote stem. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 383–412 (1993).
 - Milner, A. C. The aistopod amphibian from the Viséan of East Kirkton, West Lothian, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 363–368 (1993).
 - Jeram, A. J. & Selden, P. A. Eurypterids from the Viséan of East Kirkton, West Lothian, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 301–308 (1993).
 - Jeram, A. J. Scorpions from the Viséan of East Kirkton, West Lothian, Scotland, with a revision of the infraclass Mesoscorpionina. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 283–300 (1993).
 - Shear, W. A. Myriapodous arthropods from the Viséan of East Kirkton, West Lothian, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 309–316 (1993).
 - Wood, S. P., Panchen, A. L. & Smithson, R. R. A terrestrial fauna from the Scottish lower Carboniferous. *Nature* **314**, 355–356 (1985).
 - Clarkson, E. N. K., Milner, A. R. & Coates, M. I. Palaeoecology of the Viséan of East Kirkton, West Lothian, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 417–426 (1993).
 - Beaumont, E. H. Cranial morphology of the Loxomatidae (Amphibia: Labyrinthodontia). *Phil. Trans. R. Soc. B* **280**, 29–101 (1977).
 - Coates, M. I. The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial skeleton, basal tetrapod relationships and patterns of skeletal evolution. *Trans. R. Soc. Edinb. Earth Sci.* **87**, 363–421 (1996).
 - Carroll, R. L. & Gaskill, P. The Order Microsauria. *Mem. Am. Phil. Soc.* (1978).
 - Milner, A. C. & Lindsay, W. Postcranial remains of *Baphetes* and their bearing on the relationships of the Baphetidae (=Loxomatidae). *Zool. J. Linn. Soc.* **122**, 211–235 (1998).
 - Laurin, M. & Reisz, R. R. in *Amniote Origins—Completing the Transition to Land* (eds Sumida, S. S. & Martin, K. L. M.) 9–59 (Academic, London, 1997).
 - Holmes, R. B., Carroll, R. L. & Reisz, R. R. The first articulated skeleton of *Dendrerpeton acadianum* (Temnospondyli, Dendrerpetontidae) from Joggins (Westphalian B) Nova Scotia, and the interrelationships of early temnospondyls. *J. Vert. Paleontol.* **18**, 64–79 (1998).
 - Beaumont, E. H. & Smithson, T. R. The cranial morphology and relationships of the aberrant Carboniferous amphibian *Spathicephalus mirus* Watson. *Zool. J. Linn. Soc.* **122**, 187–209 (1998).
 - Smithson, T. R. The cranial morphology of *Greerpeton burkemorani* Romer (Amphibia: Temnospondyli). *Zool. J. Linn. Soc.* **76**, 29–90 (1982).
 - Clack, J. A. The stapes of the Coal Measures embolomere *Pholidrpeton scutigerum* Huxley and otic evolution in early tetrapods. *Zool. J. Linn. Soc.* **79**, 121–148 (1983).
 - Clack, J. A. The evolution of tetrapod ears and the fossil record. *Brain Behav. Evol.* **50**, 198–212 (1997).
 - Lebedev, O. A. & Coates, M. I. The postcranial skeleton of the Devonian tetrapod *Tulerpeton curtum* Lebedev. *Zool. J. Linn. Soc.* **114**, 307–348 (1995).
 - Swofford, D. L. *PAUP: Phylogenetic Analysis using Parsimony, Version 3.1* (Illinois Nat. Hist. Surv., Champaign, Illinois, 1993).
 - Maddison, W. P. & Maddison, D. R. *MacClade, Version 3.0 Analysis of Phylogeny and Character Evolution* (Sinauer Associates, Sunderland, MA, 1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. I thank A. C. Milner for access to unpublished work on the postcranial material of baphetids; A. C. Milner and A. R. Milner for discussion of baphetid relationships and for comments on draft manuscripts; and K. A. Joysey for facilitating the acquisition of specimens of *Eucritta*.

Correspondence and requests for materials should be addressed to the author (e-mail: jac18@hermes.cam.ac.uk).

Adaptive radiation in a heterogeneous environment

Paul B. Rainey & Michael Travisano

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

Successive adaptive radiations have played a pivotal role in the evolution of biological diversity^{1–3}. The effects of adaptive radiation are often seen^{4–6}, but the underlying causes are difficult to disentangle and remain unclear^{7–9}. Here we examine directly the role of ecological opportunity and competition in driving genetic diversification. We use the common aerobic bacterium *Pseudomonas fluorescens*¹⁰, which evolves rapidly under novel environmental conditions to generate a large repertoire of mutants^{11–13}. When provided with ecological opportunity (afforded by spatial structure), identical populations diversify morphologically, but when ecological opportunity is restricted

there is no such divergence. In spatially structured environments, the evolution of variant morphs follows a predictable sequence and we show that competition among the newly evolved niche-specialists maintains this variation. These results demonstrate that the elementary processes of mutation and selection alone are sufficient to promote rapid proliferation of new designs and support the theory that trade-offs in competitive ability drive adaptive radiation^{14,15}.

Explanation of macroevolutionary phenomena (for example, adaptive radiation and punctuated evolution) by direct extrapolation from microevolutionary processes (for example, mutation and competition) is contentious^{1,16–18}. Conventional explanations for adaptive radiation frequently invoke no more than vacant niches and stringent competition between niche specialists^{14,19–21}. Experimental studies have lent credence to this view²², but by necessity

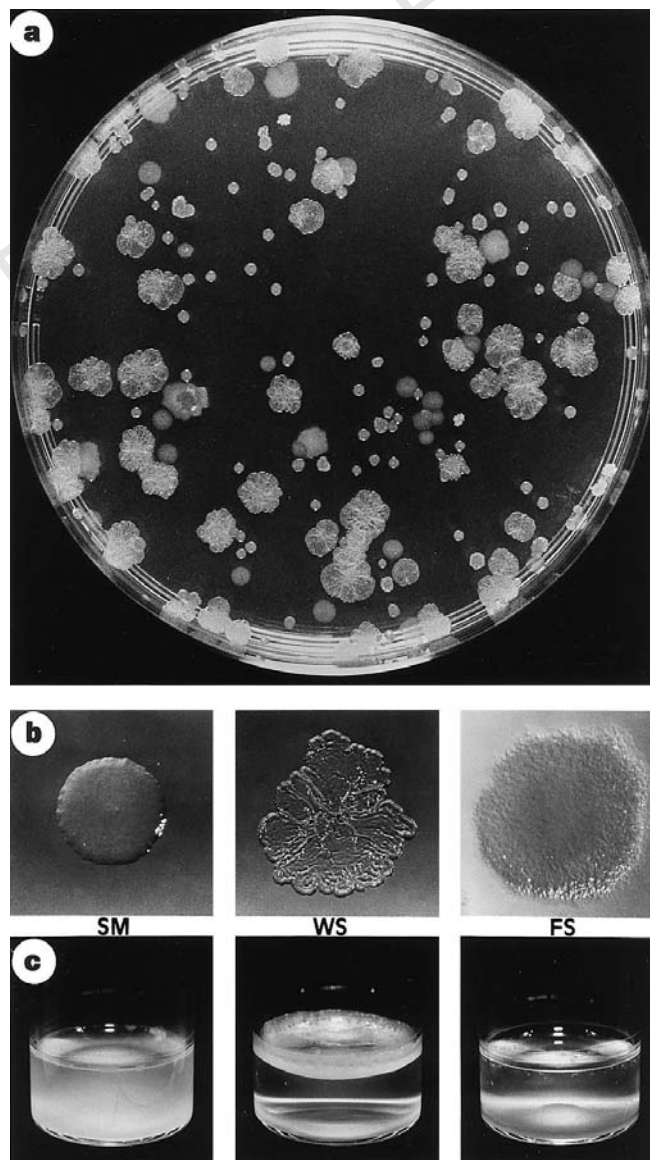


Figure 1 Phenotypic diversity and niche specificity among *P. fluorescens* SBW25 colonies evolved in a spatially heterogeneous environment. Populations were founded from single ancestral 'smooth' (SM morph) cells and propagated in 6-ml King's medium B contained in a 25-ml microcosm at 28°C. Microcosms were incubated without shaking to produce a spatially heterogeneous environment. **a**, After 7 days, populations show substantial phenotypic diversity which is seen after plating. **b**, Most phenotypic variants can be assigned to one of three principle morph classes: (SM), wrinkly spreader (WS) and fuzzy spreader (FS). **c**, Evolved morphs showed marked niche preferences.

these studies have been limited in scope^{7,15,21}. An ideal experiment would follow the evolution of a single genotype in multiple environments that differ solely in their niche potential. The experiment would be designed such that ecological and genetic factors could be carefully controlled, thus allowing mechanistic hypotheses concerning the origin and maintenance of diversity to be tested.

Bacteria are well suited for such a study. They are easily propagated, have rapid generation times and large population sizes, and can be stored in a state of suspended animation so that the fitness assays of evolved and ancestral genotypes can be performed by direct competition^{9,23}. As bacteria reproduce asexually, identical populations can be established from a single genotype; all subsequent variation is therefore generated *de novo*, by mutation. *P. fluorescens* (SBW25) populations evolve rapidly in novel environments¹². Evolution is seen at the phenotypic level in differences in colony morphology, which are easy to detect (Fig. 1), heritable and genetically based (see Methods). A striking feature of the evolved morphs is their niche specificity (Fig. 1c). The fortuitous relationship between colony morphology and niche preference allows real-time ecological and evolutionary dynamics to be determined by scoring changes in the frequencies of colony phenotypes.

To investigate the effect of ecological opportunity on the evolution of genetic diversity, we propagated replicate isogenic populations of the ancestral (smooth) morph in spatially heterogeneous environments provided by static broth cultures (microcosms; Fig. 1). Microcosms were destructively sampled every 24 h and the evolution of diversity was determined by monitoring changes in the frequencies of colony morphologies. After 3 days, extensive morphological diversification was evident and this was sustained over 10 days (Fig. 2a). The three dominant morphs (named smooth, wrinkly spreader and fuzzy spreader) showed highly repeatable evolutionary dynamics across populations (Fig. 2c). This fact and

the rapid rise in diversity indicated that strong, diversifying selection was occurring. To test whether ecological opportunity, afforded by multiple vacant niches, was a crucial factor in the evolution of diversity, we propagated replicate populations of the ancestral (smooth) morph as before, but destroyed the physical structure of the environment by shaking the cultures, thereby eliminating the spatial heterogeneity and the multiplicity of niches. In this essentially homogeneous environment, there was no morphological variation (Fig. 2b, d).

Initial diversification is not necessarily indicative of sustained diversity^{4,19} and it is possible that the differences in morphological diversity between populations in the two environmental regimes were transient. Variation in the heterogeneous environment may have resulted from sequential sweeps of adaptive mutants that would eventually leave a single dominant phenotype²⁴. Alternatively, successive selective sweeps of undetected traits may have prevented morphological diversification in the homogeneous environment. To investigate these possibilities, we propagated a set of ten isogenic ancestral (smooth) populations under the spatially heterogeneous regime. After 7 days these cultures, as expected, became genetically diverse. We then determined whether the maintenance of diversity depended on continued environmental heterogeneity. We used a sample from each of the ten genetically diverse populations to found two new populations, one set of which was evolved for two successive 7-day periods under spatially heterogeneous conditions, the other of which was evolved under spatially homogeneous conditions (Fig. 3). Populations in the heterogeneous environment showed sustained levels of high diversity, whereas populations switched to the homogeneous environment suffered a rapid loss of diversity. These results indicate that selection is important in the maintenance of morphological diversity in the heterogeneous environment, but leads to purging of variation in the homogeneous environment.

If selection is the primary evolutionary force maintaining diversity in the heterogeneous environment, then competitive trade-offs among different niche-adapted genotypes would be expected^{14,15,21,25}. Such trade-offs are likely to be frequency-dependent²⁶: that is, genotypes will have a fitness advantage when they are rare that disappears when they are common. To assess this, we examined the ability of the three principal morphological types to invade, when rare, a population of a different morph type. The

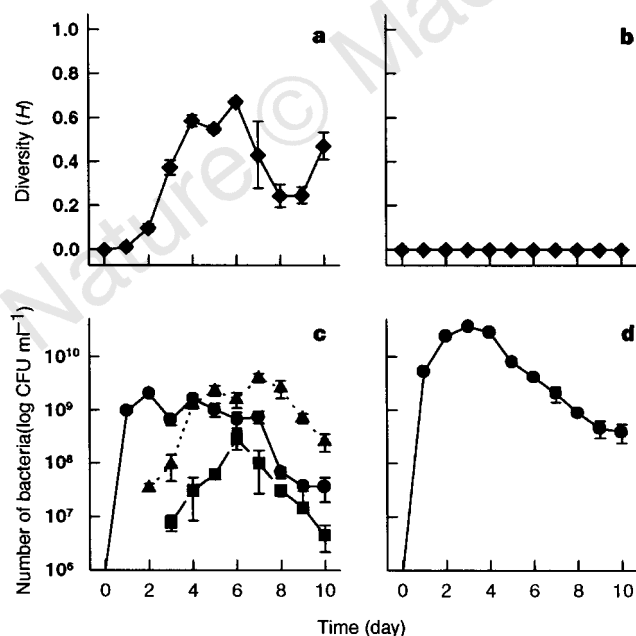


Figure 2 Effect of ecological opportunity on the evolution of genetic diversity. Two sets of identical populations were founded from the ancestral smooth morph and incubated either with or without shaking. Every 24 h, three replicate microcosms were destructively collected and diversity was determined by scoring the frequencies of morphs. **a, b**, Diversity increased rapidly in the heterogeneous environment (**a**) whereas no diversity was detected in environments lacking spatial structure (**b**). **c, d**, Evolutionary dynamics of the principal morph classes in the heterogeneous (**c**) and homogeneous (**d**) environments. Values are the means $\pm$ s.e.m. ($n = 3$). Circles represent smooth morphs, triangles represent wrinkly-spreader morphs and squares represent fuzzy-spreader morphs.

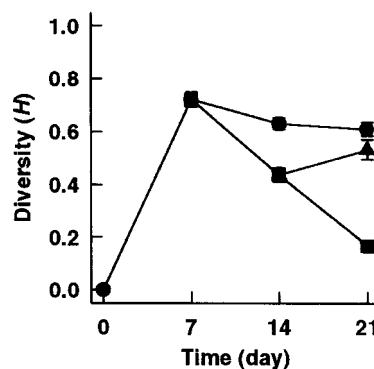


Figure 3 Effect of spatial heterogeneity on the maintenance of genetic variation. Ten replicate populations founded from the ancestral smooth morph were allowed to evolve in static microcosms. After 7 days, a sample of $\sim 4 \times 10^4$ cells from each microcosm was used to found two fresh sets of ten replicate microcosms; one set was incubated under spatially homogeneous conditions (squares) and the other under spatially heterogeneous conditions (circles; upper line). These populations were propagated for two additional 7-day periods. Diversity in homogeneous microcosms declined significantly relative to their heterogeneous partners ($P < 0.001$, one-tailed test). Propagation of day-14 homogeneous microcosms under heterogeneous conditions restored diversity (triangle). Values are the mean $\pm$ s.e.m. ($n = 10$).

single ancestral (smooth) genotype and 24 independently derived isolates from each of the wrinkly-spreader and fuzzy-spreader evolved classes were competed in a pairwise manner. The ratio of competing genotypes at the beginning of each experiment was $\sim 100:1$, and populations were competed for 7 days to determine the fitness of the initially rare genotype relative to the common genotype by scoring frequencies on agar plates. Use of a neutral pantothenate marker allowed us to distinguish between invasion of the initially rare genotype and phenotypically visible diversification of the common genotype. From three sets of pairwise competitions (smooth versus wrinkly spreader, smooth versus fuzzy spreader and fuzzy spreader versus wrinkly spreader), two interactions resulted in stable maintenance of variation (Fig. 4), namely the interactions involving the ancestor and derived morphs (smooth versus fuzzy spreader or wrinkly spreader). The detection of competitive trade-offs between niche specialists (frequency-dependent selection) confirms predictions of population genetics theory and indicates that the engine of adaptive radiation was competition^{14,25}.

The ecological mechanisms maintaining diversity within the spatially heterogeneous environment are complex and not fully understood. However, competition for oxygen seems to be a significant factor. Cells of the wrinkly-spreader morph adhere firmly to each other and to surfaces, allowing the formation of a self-supporting mat (Fig. 1). Occupation of the air–broth interface provides WS cells with access to both oxygen and nutrients and so they are selectively favoured at the expense of genotypes within the oxygen-poor broth phase. Once the wrinkly-spreader morph becomes common, its fitness declines because the weight of the mat increases beyond the point at which it is self-supporting, and it sinks. The selective advantage of the fuzzy-spreader morph is unclear, but its persistence appears to depend on the stable frequency-dependent selection between the smooth and wrinkly-spreader morphs.

It is striking that complex ecological interactions within the spatially heterogeneous environment evolved over such a short period of time, and that the interactions were nearly identical across microcosms. The rapidity and repeatability of evolution here are indicative of both strong selective pressures and the adaptive potential of the ancestral genotype, despite the underlying random nature in the appearance of mutations. However, the apparent determinism of the system is not all-encompassing. First, although we detected three dominant morphological classes

in every microcosm, variation was always encountered within the morphological classes across microcosms, and different, minor morphological classes appeared (we have not described these here). Second, in experiments performed in spatial microcosms one-hundredth the size of those described above, diversification resulted in far less predictable patterns, indicating that the apparent determinism is due, in part, to large population sizes and hence large numbers of mutants. Finally, persistence of the three dominant morphological classes indicates that multiple avenues for adaptation existed for the ancestral genotype, and shows that different cell lineages followed different adaptive pathways.

Variation is extensive within populations of many organisms and is thought to correlate with environmental heterogeneity²⁷. In bacteria, variation within a clonal lineage is commonplace^{11–13}; stable polymorphisms have been observed in experimental populations^{28,29} and variation in pathogen populations is evident in plate culture—in some instances the molecular basis for the variation is known^{12,13}. Despite intensive study, the causes and significance of this variation have remained elusive^{11–13,30}. By showing that different variants occupy different niches and that competition between variants drives diversification, we reveal both the ecological causes and the evolutionary significance of morph variation within *P. fluorescens* populations. In so doing, we provide a framework within which to consider the significance of phenotypic variation in other bacteria.

Rigorous tests of the significance of ecological opportunity in adaptive radiation have been limited¹. Here, within a matter of days we have seen an adaptive radiation that has the hallmarks of macroevolutionary dynamics, including rapid evolution and niche specialization. By examining the evolution of genetically uniform populations in identical environments that differ only in their niche potential, we have shown that spatial heterogeneity is a significant factor affecting not only the occurrence, but also the maintenance, of diversity. Multiple vacant niches provided opportunities for niche specialization and led to the partitioning of genotypes according to habitat. The driving force for this radiation was competition. As populations were founded by a single asexually reproducing genotype, we can attribute the evolution and proliferation of new designs directly to mutation and natural selection. □

Methods

Bacteria. *P. fluorescens* isolate SBW25 was isolated from a sugar beet leaf¹⁰ and propagated in King's medium B (KB). Long-term storage was at -80°C . A pantothenate auxotroph of *P. fluorescens* SBW25 was generated by deletion of the entire *panB* gene from the ancestral smooth genotype. The marker itself had no discernible effect on fitness in the environments used in this study.

Genetic basis of morph phenotype. Twenty replicate microcosms founded by the pantothenate-marked ancestral genotype (10 Pan⁺, 10 Pan[−]) were propagated for 3 days in static microcosms. A single wrinkly-spreader morph was isolated from each microcosm and grown to stationary phase. Pairs of morphs with opposing markers were mixed and plated onto KB. The correspondence between phenotype and pantothenate marker was determined by growth assays on media supplemented with and without pantothenate. Correspondence was statistically significant ($P = 5 \times 10^{-35}$).

Diversity. Diversity (H) was calculated using the Shannon–Weaver index [$H = (N \log N - \sum n_i \log n_i)/N$], where N is the total number of individuals and n_i is the number of individuals of each phenotype. The phenotypes of at least 100 randomly selected colonies were scored from each microcosm.

Competition experiments. Forty-eight replicate microcosms founded by the smooth morph (24 Pan⁺, 24 Pan[−]) were propagated for 3 and 7 days, after which time 24 wrinkly-spreader and 24 fuzzy-spreader colonies were isolated, respectively. For each of the six interactions between niche-adapted classes, each derived niche-adapted genotype was competed against the common ancestor or a randomly chosen niche-adapted genotype of the other class. We used two-tailed sign tests to assess the statistical significance of the competitions, and all were significant at $P < 0.01$, after sequential Bonferroni correction for carrying out six simultaneous tests.

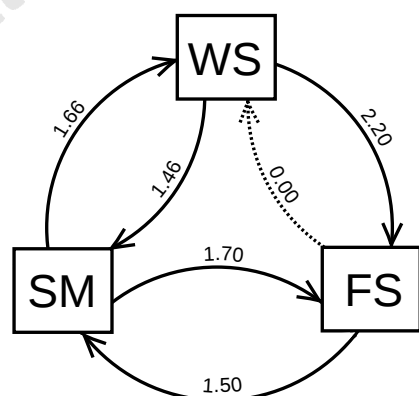


Figure 4 Competitive relationships among niche-adapted classes. Arrows point from the initially rare (invading) morph to the common (invaded) morph for each of the competitions. Values above each curve are the median fitnesses. Fitness measures were determined over 7 days by the ratio of Malthusian parameters of the initially rare genotype to the common genotype, so that a fitness of 1.00 indicates genotypes of equal competitive ability²³. In five out of six cases, the rare morph successfully invaded the common morph, whereas in one case (FS invading WS) the rare genotype was selected against. All interactions are statistically significant ($P < 0.01$) by two-tailed tests.

Received 20 October 1997; accepted 25 March 1998.

- Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* (Penguin, London, 1989).
- Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).
- Benton, M. J. Diversification and extinction in the history of life. *Science* **268**, 52–58 (1996).
- Otte, D. & Endler, J. A. (eds) *Speciation and its Consequences* (Sinauer, Sunderland, 1989).
- Dobzhansky, T. Species of *Drosophila*. *Science* **177**, 664–669 (1972).
- Klein, D. et al. Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature* **364**, 330–334 (1993).
- Schluter, D. Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**, 798–801 (1994).
- Grant, P. R. *Ecology and Evolution of Darwin's Finches* (Princeton Univ. Press, NJ, 1986).
- Lenski, R. E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000 generation experiment with bacterial populations. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814 (1994).
- Rainey, P. B. & Bailey, M. J. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol. Microbiol.* **19**, 521–533 (1996).
- Shapiro, J. A. in *The Bacteria; The Biology of Pseudomonas* (ed. Sokatch, J. R.) 27–69 (Academic, London, 1986).
- Rainey, P. B., Moxon, E. R. & Thompson, I. P. Intracolonial polymorphism in bacteria. *Adv. Microb. Ecol.* **13**, 263–300 (1993).
- Moxon, E. R., Rainey, P. B., Nowak, M. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33 (1994).
- Taper, M. T. & Case, T. J. in *Oxford Surveys in Evolutionary Biology*, Vol. 8 (eds Futuyma, D. J. & Antonovics, J.) 63–109 (Oxford Univ. Press, Oxford, 1994).
- Connell, J. H. The influence of interspecific competition and other factors on the distribution of the barnacle *Chthamalus stellatus*. *Ecology* **42**, 710–723 (1961).
- Charlesworth, B., Lande, R. & Slatkin, M. A neo-Darwinian commentary on macroevolution. *Evolution* **46**, 16–30 (1982).
- Travisano, M., Mongold, J. A., Bennett, A. F. & Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995).
- Elena, M. S. F., Cooper, V. S. & Lenski, R. E. Punctuated evolution caused by selection of rare beneficial mutations. *Science* **272**, 1802–1804 (1996).
- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
- Conway Morris, S. & Whittington, H. B. The animals of the Burgess shale. *Sci. Am.* **240**, 122–133 (1979).
- Hopf, F. A., Valone, T. J. & Brown, J. H. Competition theory and the structure of ecological communities. *Ecol. Monogr.* **7**, 142–154 (1993).
- Bell, G. & Rebut, X. Experimental evolution in *Chlamydomonas*. II. Genetic variation in strongly contrasted environments. *Heredity* **78**, 498–506 (1997).
- Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991).
- Atwood, K. C., Schneider, L. K. & Ryan, F. J. Periodic selection in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **37**, 146–155 (1951).
- Abrams, P. A. Alternative models of character displacement and niche shift. 2. Displacement when there is competition for a single resource. *Am. Nat.* **130**, 271–282 (1987).
- Haldane, J. B. S. *The Causes of Evolution* (Longmans & Green, London, 1932).
- Hedrick, P. W. Genetic polymorphism in heterogeneous environments: a decade later. *Annu. Rev. Ecol. Syst.* **17**, 535–566 (1986).
- Rosenzweig, R. E., Sharp, R. R., Treves, D. S. & Adams, J. Microbial evolution in a simple unstructured environment—genetic differentiation in *Escherichia coli*. *Genetics* **137**, 903–917 (1994).
- Turner, P. E., Souza, V. & Lenski, R. E. Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. *Ecology* **77**, 2119–2129 (1996).
- Korona, R., Nakatsu, C. H., Forney, L. J. & Lenski, R. E. Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. *Proc. Natl Acad. Sci. USA* **91**, 9037–9041 (1994).

Acknowledgements. We thank K. McCallum for technical assistance; J. Baker for photography; D. Ebert, S. Kahn, B. Haubold, and E. R. Moxon for discussion; and R. E. Lenski, P. Sniegowski and I. Moore for comments on the manuscript. This work was supported in part by grants from the B.B.S.R.C. and BTP Plc.

Correspondence and requests for materials should be addressed to P.B.R. (e-mail: prainey@worf.molbiol.ox.ac.uk).

Temporal gating of neural signals during performance of a visual discrimination task

Eyal Seidemann, Ehud Zohary* & William T. Newsome

Howard Hughes Medical Institute and Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA

* Department of Neurobiology, Institute of Life Sciences, Hebrew University, Givat Ram, Jerusalem 91904, Israel

The flow of neural signals within the cerebral cortex must be subject to multiple controls as behaviour unfolds in time. In a visual discrimination task that includes a delay period, the transmission of sensory signals to circuitry that mediates memory, decision-making and motor-planning must be governed closely by 'filtering' or 'gating' mechanisms so that extraneous

events occurring before, during or after presentation of the critical visual stimulus have little or no effect on the subject's behavioural responses. Here we study one such mechanism physiologically by applying electrical microstimulation^{1–3} to columns of directionally selective neurons in the middle temporal visual area^{4–9} at varying times during single trials of a direction-discrimination task. The behavioural effects of microstimulation varied strikingly according to the timing of delivery within the trial, indicating that signals produced by microstimulation may be subject to active 'gating'. Our results show several important features of this gating process: first, signal flow is modulated upwards on onset of the visual stimulus and downwards, typically with a slower time course, after stimulus offset; second, gating efficacy can be modified by behavioural training; and third, gating is implemented primarily downstream of the middle temporal visual area.

Several lines of evidence indicate a critical role for MT (the middle temporal visual area) in the analysis of visual motion information. Roughly 90% of MT neurons respond selectively to stimulus motion within a restricted range of directions⁷. Furthermore, MT neurons are organized in columns such that neighbouring neurons preferentially respond to similar directions of motion⁹. Electrical microstimulation of these columns can bias a monkey's judgements of motion direction towards the direction encoded by the stimulated neurons, showing that directional signals in MT are important in generating psychophysical performance^{1–3}. In earlier microstimulation experiments, trains of stimulating pulses were delivered at the same time as presentation of the motion stimulus to

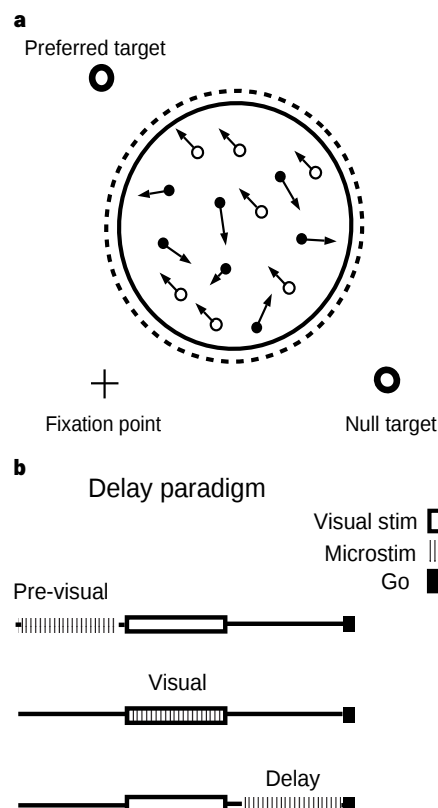


Figure 1 The protocol. **a**, Diagram of the visual display. The visual stimulus aperture (complete circle) was positioned within the multi-unit receptive field (dashed circle) mapped at the microstimulation site. Open dots represent the coherent motion signal (arrows directed upwards and left); solid dots represent the masking motion noise (randomly directed arrows). **b**, The sequence of events. Stimulating pulses were delivered for equal amounts of time in each experimental condition. Control trials with no microstimulation were randomly interleaved among the test trials.

be discriminated by the monkey. We now compare the efficacy of microstimulation delivered before, during and after presentation of the visual stimulus.

We trained four rhesus monkeys to discriminate between opposite directions of motion in a dynamic random dot display (Fig. 1a). In each experiment, a microelectrode was positioned within a column of directionally selective MT neurons. The monkey then performed the discrimination task for a block of several hundred trials in which the direction and the coherence of the motion signal varied randomly from trial to trial. Microstimulation was applied during the stimulus presentation interval or during the delay period for monkey K, and also during a fixation interval before the onset of the visual stimulus for the other three monkeys (Fig. 1b).

Figure 2 shows the results of a representative experiment from each of two of the monkeys. The psychometric functions show, for each motion condition tested, the proportion of trials in which the monkey chose the direction preferred by neurons at the stimulation site. Microstimulation during the visual stimulus interval caused both monkeys to make substantially more decisions favouring the preferred direction, resulting in a large leftward shift of the psychometric function relative to the control function (22.6% and 19.5% coherence in Fig. 2a, b, respectively; see Methods). Microstimula-

tion during the delay period influenced both monkeys, but to different degrees: the delay-period effect was substantially smaller than the visual-period effect in monkey K (8.4% coherence), but the effects were equivalent in monkey R (26.4% coherence). In contrast to the large effects in the delay period, monkey R's choices were completely unaffected by microstimulation delivered before the onset of the visual stimulus (Fig. 2b), consistent with our previous results¹.

Figure 3 shows average results from each of the four monkeys. In monkeys K and T, delay-period effects were much smaller than visual-stimulus-period effects (Fig. 3a, b; paired *t*-test, $P < 0.005$ for both monkeys). In monkeys R and S, on the other hand, delay-period effects were roughly equal to visual-period effects (Fig. 3c, d; paired *t*-test, $P > 0.5$ for both monkeys). Stimulation before the visual-stimulus interval was ineffective in monkey R, and only weakly effective in monkeys T and S.

The data indicate that the behavioural efficacy of microstimulation is modulated upwards at the time of stimulus onset, and, in monkeys K and T, is modulated downwards again at stimulus offset. The large delay-period effect in monkeys R and S was quite surprising as no visual stimulus was present during this interval. One could explain this result if some MT neurons are active during the delay period, signalling the remembered direction of motion. In this case, microstimulation during the delay period could influence behaviour by modifying the ongoing representation of past stimuli. We did not, however, observe such delay-period activity in single-unit recordings from MT neurons in either monkey R ($n = 21$ cells) or monkey K ($n = 16$ cells; E.S. and W.T.N., unpublished observations). We therefore wondered whether the efficacy of the pathway linking MT and premotor circuitry might also be modulated downwards during the delay period in monkeys R and S, but over a longer time course than in monkeys K and T.

To test this possibility, we carried out additional experiments in monkey R, applying microstimulation after increasingly longer delays following the offset of the visual stimulus. Figure 4 shows that the effect of microstimulation indeed waned substantially as the stimulating pulses were progressively delayed relative to the offset of the visual stimulus. Averaged across ten experiments, microstimulation applied after the longest delay was only 32% as effective as

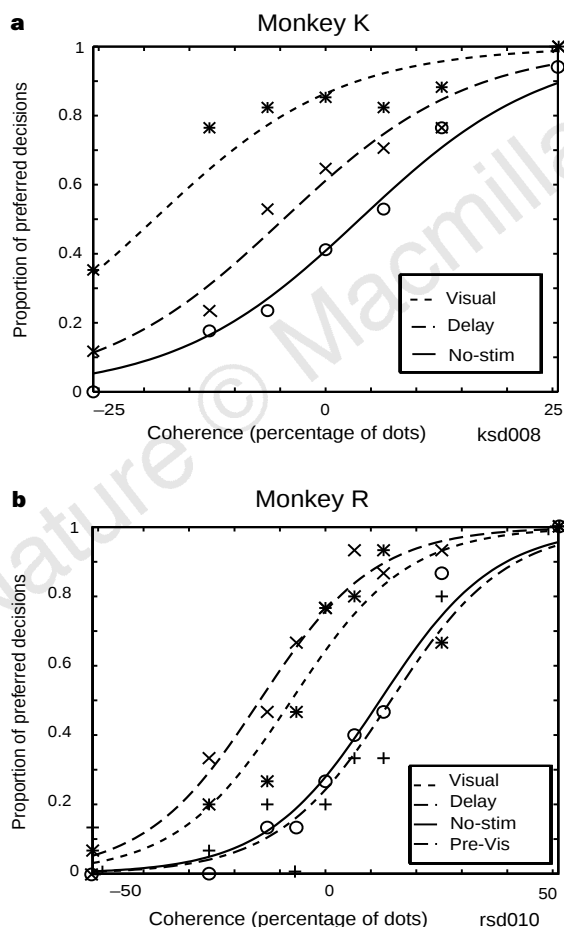


Figure 2 Data from representative experiments in monkeys K and R. **a**, Results from monkey K; and **b**, from monkey R. Each data point indicates the proportion of the monkey's decisions that favoured the preferred direction of the stimulated neurons (ordinate) as a function of stimulus coherence and direction (abscissa). Stimulus motion in the preferred direction is indicated by positive coherence values; motion in the null direction is indicated by negative values. The symbol for each data point indicates the control or test condition: no stimulation (open circles); microstimulation before (plus symbols), during (asterisks) and after (crosses) presentation of the visual stimulus. Sigmoids illustrate the fitted curve for each condition.

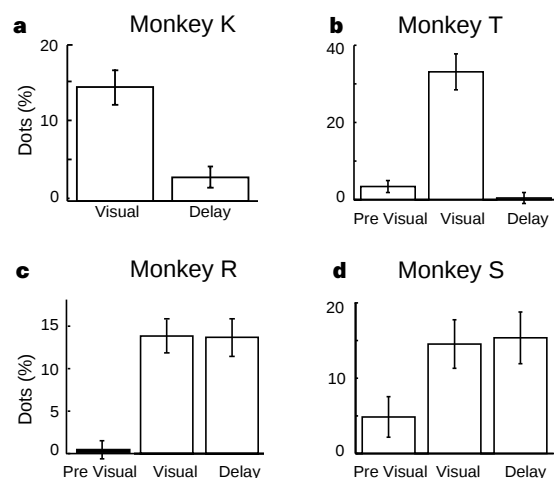


Figure 3 Average microstimulation effects in the four monkeys. Microstimulation effects (ordinate) are given as the leftward shift of the psychometric function in per cent coherent dots. For an experiment to be included in this analysis, microstimulation had to generate a significant positive shift of the psychometric function in at least one test condition (logistic regression analysis, $P < 0.05$). Negative microstimulation effects were observed in only four experiments and were excluded from this analysis. Error bars indicate the standard error of the mean. **a**, Monkey K ($n = 12$). **b**, Monkey T ($n = 13$). **c**, Monkey R ($n = 13$). **d**, Monkey S ($n = 9$).

microstimulation delivered after a short delay. This decrease in microstimulation efficacy is not simply due to the lengthening of the delay. Microstimulation effects remained high in other test conditions, randomly interleaved among those in Fig. 4, in which microstimulation was applied only during the first 900-ms interval after offset of the visual stimulus (mean shift is 15.5, 11.8 and 14.3% coherence for short, medium and long delays, respectively). Thus the behavioural efficacy of microstimulation is indeed modulated downwards in monkey R after offset of the visual stimulus, the time course of modulation being longer than in monkeys K and T.

Interanimal differences in the time course of the downward modulation were unexpectedly large, and we wondered whether the differences might be reduced if monkeys R and S were trained explicitly to ignore irrelevant information arriving during the delay period. To test this, we constructed a new behavioural task in which an irrelevant visual stimulus appeared during the delay in place of the microstimulation pulses. The first time period contained a red random dot stimulus that carried the signal for the direction discrimination; the second period contained a green random dot stimulus that was irrelevant to the discrimination. After the animals learned to ignore the green dots effectively, we repeated the standard microstimulation experiment outlined in Fig. 1, which contained a single visual stimulus period. In both monkeys, after being trained with the red–green paradigm, microstimulation during the delay period exerted a substantially weaker effect relative to stimulation during the visual stimulus period (compare Figs 3c, d and 5a, b). Before training on the red–green task, the efficacy of microstimulation in the delay period had remained roughly constant over the course of several months. Thus the change in efficacy shown in Fig. 5a, b can be attributed unambiguously to training with the red–green task. The reduced efficacy of delay-period microstimulation shows that past experience can strongly affect the gating process that

prevents irrelevant neural signals from influencing premotor circuitry. These results also indicate that the neuronal signals that are produced in MT by visual stimuli and by electrical microstimulation are subject to the same gating mechanism.

Finally, we recorded single-unit activity in MT while monkeys R and T performed the red–green task in order to determine whether the gating effect occurs at the level of MT. Weaker responses to the green (ignored) dots than to the red (attended) dots would indicate that the gating mechanism is probably expressed in MT, whereas equivalent responses to the red and the green dots would indicate that the gating operation is implemented downstream of MT. Across 42 MT cells from monkeys R and T, responses were only slightly stronger to the red dots than to the green dots (8% difference). These effects are an order of magnitude smaller than attentional effects recently observed in the MT¹⁰. In our task, therefore, the gating mechanism that prevents irrelevant stimuli from affecting psychophysical decisions seems to reside primarily downstream from MT.

The behavioural efficacy of microstimulation in MT is temporally modulated during single trials, indicating that an active gating mechanism controls information flow from MT to premotor circuitry. Microstimulation allowed us to study the time course of the gating process; signal flow was modulated upwards rapidly at stimulus onset and downwards, with a variable time course, at stimulus offset. Using microstimulation as a probe, we found that the time course of downward modulation can be affected strongly by the monkey's past experience. The gating mechanism operates primarily downstream of MT. Our data place no constraints on the length of the pathways that link MT to precolumotor circuitry; we can only say that the gating mechanism acts at some point on these pathways, whatever their length. Similarly, our experiments place no constraint on the cellular mechanisms underlying the gating process: facilitation of relevant signals and inhibition of irrelevant signals (or both) are equally plausible.

Gating of information flow according to its behavioural significance must be ubiquitous within the central nervous system; its existence can be inferred from numerous psychophysical and physiological studies, including those on the effects of visual attention (reviewed in refs 11, 12). Thus our observations concerning temporal gating are related to earlier studies that show modulation of neural signals by spatial- or feature-based attentional mechanisms^{13–17}. An important outcome of our study is the establishment of a physiological paradigm by which a neuronal signal that is subject to gating can be delivered at will by experimenters to a precise location in the central nervous system. An important goal for the future will be to follow this signal from MT to premotor circuitry in order to localize the site at which irrelevant signals are filtered out. Identification of such a site may allow cellular analysis of the gating mechanism and circuit-level analysis of the higher-level signals that control the operation of the gate. □

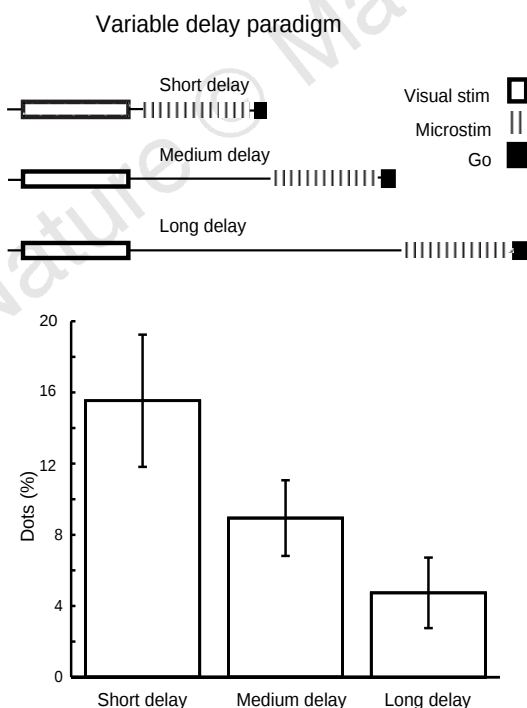


Figure 4 Sequence of events and average microstimulation effects in the variable delay paradigm. For each test condition, the amplitude of the microstimulation effect was considered to be the shift of the test psychometric function relative to a control function (no microstimulation) obtained using a matched delay period. The microstimulation effect decreased significantly as the delay period lengthened (one-way analysis of variance, $P < 0.05$). Error bars denote the standard error of the mean.

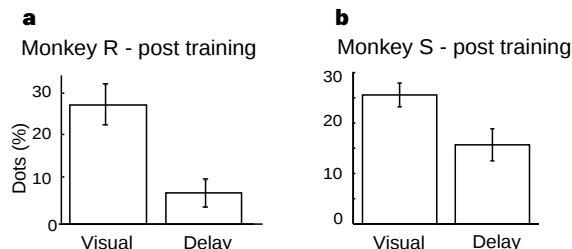


Figure 5 Average microstimulation effect in monkeys R and S following training on the red–green task. In contrast to the results shown in Fig. 3c and d, delay-period stimulation caused significantly weaker effects than microstimulation during the visual stimulus (paired t -test, $P < 0.005$ for both monkeys). **a**, Monkey R ($n = 14$). **b**, Monkey S ($n = 12$).

Methods

The treatment of the monkeys was in accordance with the guidelines set by the US Department of Health and Human Services (NIH) for the care and use of laboratory animals.

Behavioural tasks. The behavioural task and physiological techniques used here have been described¹. Briefly, in each trial, a fraction of the dots moved coherently in one direction while the remaining dots appeared at random locations within the stimulus aperture, creating a masking motion noise. The strength of the coherent motion (per cent coherence) ranged from 0% (fully random noise) to 51.2%. Each trial began when the monkey fixated a small point of light on a video display (Fig. 1a). The monkey then viewed the random dot display for a brief interval while gazing steadily at the fixation point. Following a brief delay period, the monkey indicated the perceived direction of motion by making a saccadic eye movement to one of two visual targets corresponding to the two possible directions of motion. Extinction of the fixation point and appearance of the two visual targets signalled the end of the delay period. Eye movements were measured throughout all experiments by the scleral search coil technique¹⁸. The monkey received fluid rewards for correct choices.

Microstimulation techniques. For each microstimulation experiment, the multi-unit receptive field was mapped and the preferred direction identified using a bar of light or a patch of moving dots. To maximize the likelihood that the monkey would use directional information supplied by neurons at the stimulation site, the dot patterns were presented within the receptive field, and the axis of the directional discrimination was aligned with the preferred-null axis of the receptive field. Microstimulation (200 Hz, 10 μ A, biphasic pulses) was applied during the stimulus presentation interval or during the delay interval for monkey K, and also during a fixation interval before the onset of the visual stimulus ('previsual') for the other three monkeys (Fig. 1b). For monkeys R, S and T, each microstimulation interval lasted 800 ms; 200 ms separated stimulation intervals. For monkey K, microstimulation duration was 1 s with 500 ms separation between the two microstimulation intervals. In an extra ten experiments, monkey R was tested with same time course as monkey K; the results remained essentially the same as in Fig. 3c. The latter time course was also used in the red-green experiment. In the variable delay experiment (Fig. 4), microstimulation was applied in the last 800 ms of the delay period; the duration of the delay period was varied randomly among three values—900 ms (short), 1,800 ms (medium) and 2,700 ms (long).

Data analysis. Data were analysed using a standard logistic regression model¹⁹. Microstimulation effects were measured as the horizontal shift of the psychometric function in stimulated trials relative to control trials. The amplitude of this shift was measured in units of the abscissa (coherence, in percentage of dots), where leftward shifts were considered to be positive. Throughout this study, microstimulation did not significantly affect the slopes of the psychometric functions.

Received 20 October 1997; accepted 17 April 1998.

- Salzman, C. D., Murasugi, C. M., Britten, K. H. & Newsome, W. T. Microstimulation in visual area MT: effects on direction discrimination performance. *J. Neurosci.* **12**, 2331–2355 (1992).
- Murasugi, C. M., Salzman, C. D. & Newsome, W. T. Microstimulation in visual area MT: effects of varying pulse amplitude and frequency. *J. Neurosci.* **13**, 1719–1729 (1993).
- Salzman, C. D. & Newsome, W. T. Neural mechanisms for forming a perceptual decision. *Science* **264**, 231–237 (1994).
- Dubner, R. & Zeki, S. M. Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus. *Brain Res.* **35**, 528–532 (1971).
- Baker, J. F., Petersen, S. E., Newsome, W. T. & Allman, J. M. Visual response properties of neurons in four extrastriate visual areas of the owl monkey (*Aotus trivirgatus*). *J. Neurophysiol.* **45**, 397–416 (1981).
- Van Essen, D. C., Maunsell, J. H. R. & Bixby, J. L. The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic representation. *J. Comp. Neurol.* **199**, 293–326 (1981).
- Maunsell, J. H. R. & Van Essen, D. C. Functional properties of neurons in the middle temporal visual area (MT) of the macaque monkey: I. Selectivity for stimulus direction, speed and orientation. *J. Neurophysiol.* **49**, 1127–1147 (1983).
- Albright, T. D. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* **52**, 1106–1130 (1984).
- Albright, T. D., Desimone, R. & Gross, C. G. Columnar organization of directionally selective cells in visual area MT of macaques. *J. Neurophysiol.* **51**, 16–31 (1984).
- Treue, S. & Maunsell, J. H. R. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
- Posner, M. & Petersen, S. The attention system of the human brain. *Annu. Rev. Neurosci.* **13**, 25–42 (1990).
- Desimone, R. & Duncan, J. Neural correlates of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).

- Moran, J. & Desimone, R. Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784 (1985).
- Motter, B. C. Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J. Neurosci.* **14**, 2190–2199 (1994).
- Mountcastle, V. B., Motter, B. C., Steinmetz, M. A. & Duffy, J. In *Dynamic Aspects of Neurocortical Function* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 159–193 (Wiley, New York, 1984).
- Corbetta, M., Miezin, F., Dobmeyer, S., Shulman, G. & Petersen, S. Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *J. Neurosci.* **11**, 2383–2402 (1991).
- Corbetta, M., Miezin, F., Shulman, G. & Petersen, S. A PET study of visuospatial attention. *J. Neurosci.* **13**, 1202–1226 (1993).
- Robinson, D. A. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* **10**, 137 (1963).
- Cox, D. R. *Analysis of Binary Data* (Methuen, London, 1970).

Acknowledgements. We thank J. Groh, G. DeAngelis, G. Horwitz, D. Chen and J. Nichols for criticism of the manuscript, and J. Stein for technical assistance. We were supported by the National Eye Institute. W.T.N. is an investigator of the Howard Hughes Medical Institute.

Correspondence should be addressed to E.S. (e-mail: eyal@monkeybiz.stanford.edu).

Seeing only the right half of the forest but cutting down all the trees?

Fabrizio Doricchi^{††} & Chiara Inccocchia^{*}

^{*} Centro Ricerche di Neuropsicologia, IRCCS S. Lucia, Via Ardeatina 306, 00179 Roma, Italy

^{††} Istituto di Psicologia "Luigi Meschieri", Università degli Studi di Urbino, Via Saffi 15, 61029, Urbino, Italy

Unilateral neglect following damage to the right hemisphere of the brain can be characterized by failure of the global attentional mechanisms of the right hemisphere to direct the local detail processors of the left hemisphere towards the contralesional left hemispace. This is suggested by patients who recognize the global form of the left side of shapes (the forest) but fail to cancel out its local details (the trees)¹. Here we report the opposite behavioural dissociation in a patient (Q.M.) with damage to the right hemisphere of the brain. Q.M. detected local details (such as the tail of a dog) on the left or right side of visual shapes, regardless of whether these details belonged to predefined target shapes (a dog in this case) or to distractor shapes differing on the opposite side (a dog with a swan's neck and head, for example). Psychological testing showed an abnormal tendency of this patient to respond to local features, but perfect accuracy in interpreting global features when the local features could not interfere in global processing. The results indicate that the left hemisphere can integrate multiple local features simultaneously but loses global awareness as soon as local features individually compete for response selection. However, awareness of the whole is not necessary for the sequential processing of the parts.

Q.M. is a 57-year-old right-handed woman who suffered right-hemisphere ischaemia (Fig. 1). Two months after the stroke, the patient had severe contralesional neglect and hemianopia. We examined this patient extensively in light of the following observations. When presented with an array of 3s and Bs (Fig. 2a, test a1) and required to cancel out the 3s, Q.M. cancelled out both 3s and Bs on the right side of the array, disregarding the presence of the vertical bar on the left side of the Bs. Tilting the items and varying the thickness of the vertical bar (Fig. 2a, tests a2, a3) did not modify the patient's performance. Unexpectedly, the same behaviour was observed when the items were mirror-reversed about the vertical axis so that the difference between the two stimuli was on the right, not left, side. These findings were replicated using 6s and Ss (Fig. 2a, test b). When asked about her strategy, the patient answered that she looked for items with 'two little humps' when cancelling out Bs, and for items with an upper 'hook' when cancelling out 6s.

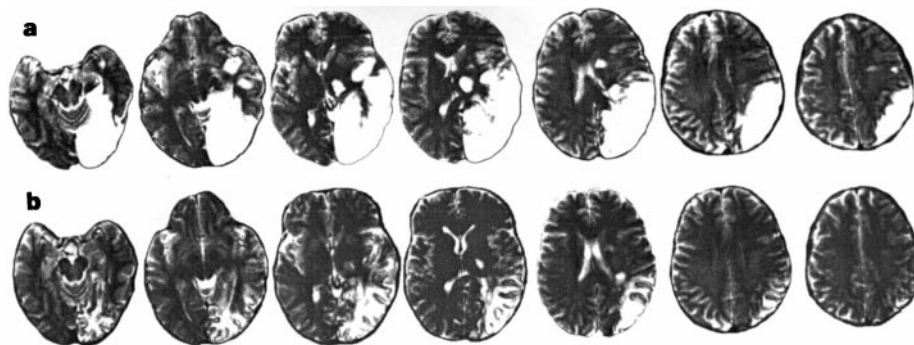


Figure 1 NMR scans of patient Q.M. Scans were taken at **a**, 5 days, and **b**, 15 days post-stroke. The right hemisphere lesion involves the occipital, temporal and parietal lobes (Brodmann's areas 17, 18, 19, 22, 37, 39 and 40) and the thalamic-capsular area. The patient's scores on a battery of tests¹² for assessment of

neglect were the following: line cancellation, 6 out of 21; letter cancellation, 14 out of 104; Wundt-Jastrow area illusion, 0 out of 20; sentence reading; 0 out of 6. Higher scores indicate better performance.

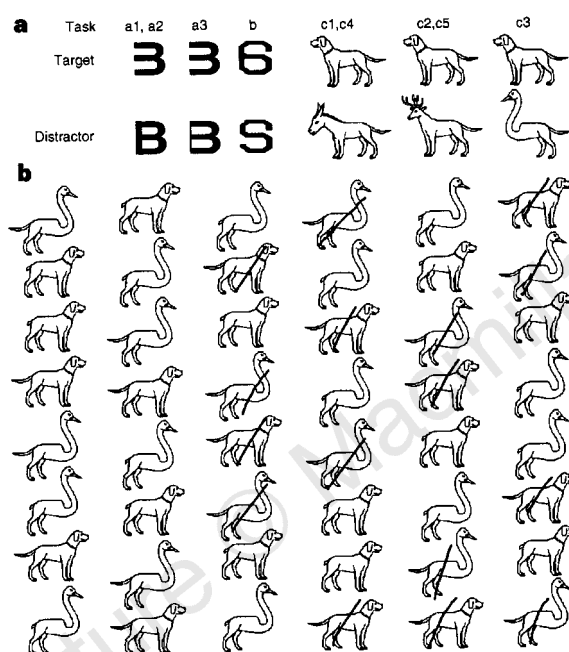


Figure 2 Cancellation tasks. **a**, Items of cancellation tasks a1-a3, b, and c1-c5. **b**, Example of Q.M.'s performance on task c3 (rightward-heading items). The ratios of frequencies of total cancelled targets to total cancelled distractors in the canonically oriented (can), mirror-reversed (mrev), leftward-heading (lhead) and rightward-heading (rhead) conditions were as follows: task a1 45:45 (can), 45:45 (mrev); task a2 39:39 (can), 42:42 (mrev); task a3 36:36 (can), 42:42 (mrev); task b 28:30 (can), 48:48 (mrev); task c1 16:16 (lhead), 16:16 (rhead); task c2 16:16 (lhead), 12:12 (rhead); task c3 16:16 (lhead), 18:18 (rhead); task c4 20:20 (lhead), 16:20 (rhead); task c5 16:10 (lhead), 16:15 (rhead).

Initially, the patient's behaviour might have been attributed to neglect based on object-centred coordinates²: in this case whatever the orientation of targets and distractors, which were items endowed with a canonical left-right orientation, the patient would have paid attention only their canonical right side, neglecting the left one. This hypothesis was ruled out by the finding that she treated items having no canonical left-right orientation in the same way that she treated 3s and 6s. When required to cancel out the line drawings of dogs with elongated tails (intermixed with short-tailed items), Q.M. omitted items on the left side of the array but cancelled out both long-tailed dogs and distractors such as a dog with a donkey's head and neck, a dog with a deer's head, or a chimera consisting of half-dog and a half-swan (Fig. 2a, b, tests c1-c3). In

searching for critical detail (that is, the elongated tail), the patient disregarded the opposite half of the animal shape. This happened independently of the leftward or rightward orientation of the items and the type of array (rows by columns, tests c1-c3; scattered items, tests c4, c5). Q.M.'s performance could not be explained by the inability to recognize living items, as she correctly named (15 out of 15) line drawings of animals³. Object-centred neglect was not found in Q.M. when copying a multi-item scene⁴, in a shape-comparison task⁵ or when reading mirror-reversed words⁶.

These initial observations indicate the presence of both neglect for the contralesional space and exacerbated narrowing of attention towards local visual details. A pathological bias towards local processing was demonstrated by tachistoscopic testing with hierarchical Navon letters⁷ (Fig. 3a) presented in the right hemifield. When Q.M. had to pay attention both to local and global features, she showed a severe deficit in detecting global ones (divided attention task, Fig. 3c). The patient suffered from exaggerated local-on-global interference and slowing of global processing even when asked to actively disregard local features (global-directed attention task, Fig. 3b). However, Q.M. recognized with perfect accuracy the global shape of stimuli in which neutral local small rectangles could neither help nor interfere with the recognition of the global letter (Fig. 3a, b). This demonstrated the retained simultaneous integration of multiple local elements in the attended ipsilesional space when these elements could not compete for response selection. Lack of interference from local rectangles following damage of the right hemisphere could also have been related to the possible involvement of this hemisphere in the processing of local features based on non-living objects⁸.

We evaluated the processing of global and local visual features in the neglected space by presenting Q.M. with symmetrical and asymmetrical hierarchical shapes (Fig. 4a). Q.M. cancelled out all the local elements of all items, scanning each element by eye and head movements with the spared visual hemifield, but she never acknowledged the different global form of the left side of asymmetrical shapes. Q.M. was also asked to mark the centre of symmetrical and asymmetrical hierarchical shapes having a longer or shorter leftward extension (Fig. 4b). The centre of symmetrical shapes was marked with a negligible rightward shift from the true centre. This could have suggested effective analysis of global shape on both sides of space. However, the centre of the asymmetrical shapes was marked extremely close to the true centre of the symmetrical ones. This showed that the influence of the global contralesional extension of shapes was dramatically small. The centre of symmetrical and asymmetrical lines lacking any vertical component favouring the enlargement of the attentional focus (Fig. 4c) was marked with a severe rightward shift, irrespective of the leftward extension of the line.

Q.M.'s case shows that left unilateral neglect together with upregulated processing of local features does not necessarily impair the direction of attention towards the contralesional local components of hierarchical stimuli. Q.M.'s lesion spared the frontal areas in the damaged hemisphere. Therefore, the sequential detection of local elements might have been based on the access of the

posterior local processors of the left hemisphere to anterior pre-motor and motor control mechanisms in both hemispheres. On the other hand, Q.M.'s right hemisphere lesion destroyed the entire network of parietal, temporal and occipital areas activated by global processing in humans^{9,10}. This finding and the associated maintenance of integration of multiple local features in the ipsilesional space support the theory¹¹ that, in patients with left unilateral neglect, global processing can still be mediated by the intact left hemisphere. However, the cancellation and tachistoscopic tasks indicated that Q.M. was prone to the loss of conscious global attention as soon as local features competed for response selection. This result specifies, for the first time to our knowledge, a sharp functional limit of the left hemisphere in global processing, and provides a new view of the assignment of local and global processing to merely relative left and right hemispheric specialization^{9,10}.

Findings from positron emission tomography⁹ show that shifting attention between local and global features depends on the influence of parietotemporal areas on prestriate visual cortex. Our observations raise a related problem, namely, how to define mechanisms allowing the maintenance of conscious awareness at one attentional level, local or global, while processing the opposite

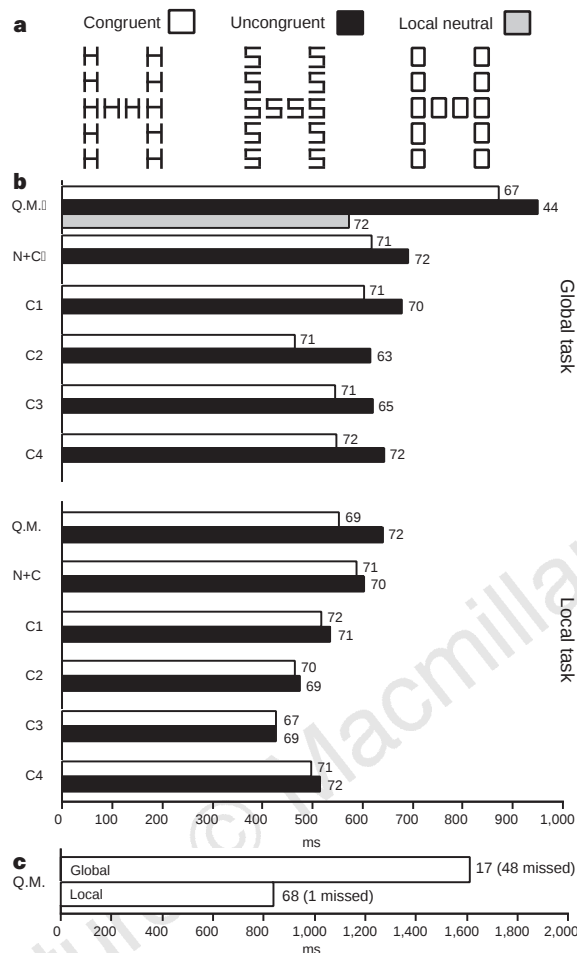


Figure 3 Bias towards local processing in patient Q.M. **a**, Examples of hierarchical Navon letters. **b**, Mean manual reaction times (in milliseconds) and frequencies of correct responses (in 72 trials; shown as numbers next to bars) of Q.M., of a control neglect patient (N + C) and of age-matched normal controls (C1–C4). In the direct attention task, Q.M. was slower (F-test $P < 0.001$) and less accurate (χ^2 test, $P < 0.01$) in the global than in the local task. In the global task: first, with incongruent stimuli Q.M.'s error frequency (in responding to local rather than global features, that is, local-on-global interference) was higher than that of the control subject (C2) with the worst performance (χ^2 test, $P < 0.001$); second, with congruent stimuli Q.M.'s performance was at ceiling (χ^2 test with continuity correction, $P > 0.5$); third, the perfect performance of Q.M. in the global task with local neutral features indicates that accuracy with congruent stimuli was probably not due to the processing of local, rather than global, features; and fourth, Q.M.'s reaction times were slower than the slowest control subject (N + C) (F-test, $P < 0.001$). In the local task: first, Q.M.'s performance did not differ from that of any control subject (χ^2 test, not significant in each comparison); and second, Q.M.'s reaction times were not statistically different from those of N + C and C1, and were slower than those of C2, C3 and C4 (F-test, $P < 0.001$ in each case). N + C (who correctly discriminated targets in the cancellation tasks with 3s, 6s and animals) had a parietal-frontal lesion which spared the posterior areas that were damaged in Q.M. **c**, Reaction times and frequencies of hits (and missed responses) in the divided attention tasks for Q.M. The detection of global targets was worse (χ^2 test, $P < 0.0001$) and slower (F-test, $P < 0.001$) than the detection of local ones.

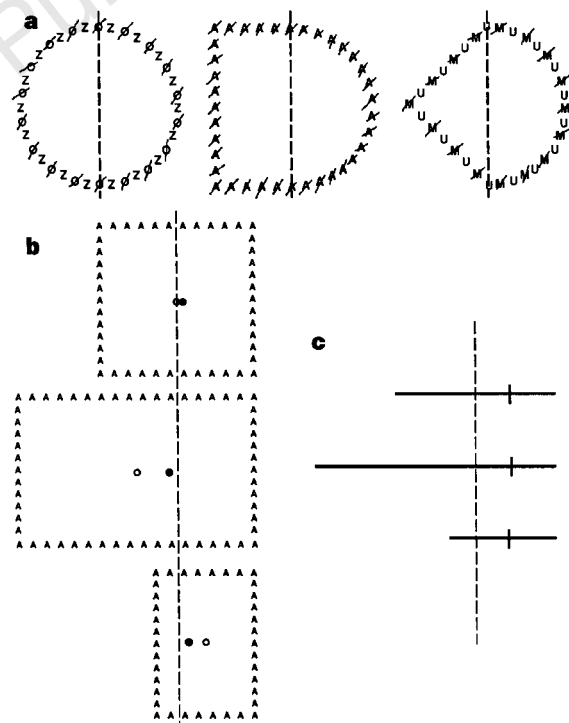


Figure 4 Processing of local and global visual features in the neglected space. **a**, Examples of cancellation by Q.M. of local elements of symmetrical and asymmetrical hierarchical shapes. **b**, Mean centre marking of hierarchical shapes. Open dot indicates true centre; filled dot indicates subjective centre. Taking the respective true centres as reference points, subjective centres were marked as follows: symmetrical shapes 2.8 mm rightward (t-test, $P < 0.01$), asymmetrical longer shapes 41.3 mm rightward (t-test, $P < 0.001$), asymmetrical shorter shapes 22.8 mm leftward (t-test, $P < 0.001$), asymmetrical shorter shapes 22.8 mm leftward (t-test, $P < 0.001$). Taking the true centres of symmetrical shapes as reference points, subjective centres were marked as follows: asymmetrical longer shapes 3.7 mm leftward, symmetrical shapes 2.8 mm rightward, asymmetrical shorter shapes 7.2 mm rightward (F omnibus test, $P < 0.001$). **c**, Mean rightward bisection error of asymmetrical longer, symmetrical, and asymmetrical shorter lines were, respectively, 39 mm, 40 mm and 37 mm (omnibus F test, not significant). The vertical dashed lines in **a–c** emphasize the relative position of items with respect to the patient's trunk midsagittal plane.

level. Q.M. exhibited a sort of transitory global simultanagnosia when attention was reflexively captured by local details which were relevant, but not sufficient, for task solution. Q.M.'s failure consciously to perceive the left half of the forest does not preclude her from cutting down all the trees one by one, but does not prevent her from failing to realize that she is cutting down the wrong woods. □

Methods

On cancellation tasks a1–a3 and b (Fig. 2), the items (2×2 cm) were arranged in an array of 9 rows $\times$ 10 columns. In tasks a2 and a3, items were tiled 45° and 90° clockwise and anticlockwise. In tasks c1–c3, items (4×4 cm) were arranged in a 8×6 array. In tasks c4 and c5 the items were scattered. Each task was performed twice for each target/distractor and item orientation condition. A total of 60 targets and 60 distractors was presented in each orientation condition of tasks a and b; 24 targets and 24 distractors were presented in each condition of tasks c.

In the tachistoscopic Navon tests, stimuli (Fig. 3a) were large letters (global features, 4×6 degrees of visual angle) composed of smaller ones (local features, 0.8×1.2 degrees). In the divided attention task (Fig. 3c), stimuli were obtained by combining, between local and global levels, letters H, S, A and E; the same letter never appeared simultaneously at both levels. The patient had to detect H and S at whatever level they appeared. In the global- and local-directed attention tasks (Fig. 3b), congruent items were letters H and S composed of smaller letters H and S and incongruent ones were letters H and S composed of smaller letters S and H, respectively. In the global task, subjects discriminated between the large letter configurations (H or S); in the local task, the small letters were discriminated between. In the global task with local neutral features, stimuli were large letters H or S composed of small rectangles. Items were presented for 150 ms, with their inner edge at 1.5° to the right of central fixation.

The hierarchical stimuli of the recognition–cancellation task (Fig. 4a) were geometrical shapes (180×180 mm) composed of small letters (5×5 mm). Symmetrical stimuli were a circle ($n = 4$), a square ($n = 4$) and a diamond ($n = 4$). Asymmetrical stimuli ($n = 12$) were obtained by combining the half shapes of the symmetrical ones. In half of the trials, a target letter was alternated with a non-target letter.

In the centre-marking task (Fig. 4b), the items were symmetrical Navon squares (180×180 mm, $n = 12$) and asymmetrical Navon rectangles with longer (270 mm, $n = 12$) or shorter (120 mm, $n = 12$) horizontal leftward extensions. The items of the bisection task (Fig. 4c) were symmetrical (180 mm, $n = 12$) and asymmetrical (270 mm, $n = 12$; 120 mm, $n = 12$) lines.

Received 16 December 1997; accepted 14 April 1998.

1. Marshall, J. C. & Halligan, P. W. Seeing the forest but only half the trees? *Nature* **379**, 521–523 (1995).
2. Walker, R. Spatial and object based neglect. *Neurocase* **1**, 371–383 (1995).
3. Snodgrass, J. G. & Vanderwart, M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* **6**, 174–215 (1980).
4. Gainotti, G., Messerli, P. & Tisot, R. Qualitative analysis of unilateral spatial neglect in relation to laterality of cerebral lesion. *J. Neurol. Neurosurg. Psychiatry* **35**, 545–550 (1972).
5. Driver, J. & Halligan, P. W. Can visual neglect operate in object centred coordinates? An affirmative single case study. *Cogn. Neuropsychol.* **8**, 475–496 (1991).
6. Hillis, A. E. & Caramazza, A. A framework for interpreting distinct patterns of hemispatial neglect. *Neurocase* **1**, 189–207 (1995).
7. Navon, D. Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* **9**, 353–383 (1977).
8. Fink, G. R. et al. Hemispheric specialization for global and local processing: the effect of stimulus category. *Proc. R. Soc. Lond. B* **264**, 487–494 (1997).
9. Fink, G. R. et al. Where in the brain does visual attention select the forest and the trees? *Nature* **382**, 626–628 (1996).
10. Robertson, L. C., Lamb, M. R. & Knight, R. T. Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *J. Neurosci.* **8**, 3757–3769 (1988).
11. Halligan, P. W. & Marshall, J. C. Focal and global attention modulate the expression of visuo-spatial neglect: a case study. *Neuropsychologia* **32**, 13–21 (1994).
12. Pizzamiglio, L. et al. Cognitive rehabilitation of the hemineglect disorder in chronic patients with unilateral brain damage. *J. Clin. Exp. Neuropsychol.* **14**, 901–923 (1992).

Acknowledgements. We thank Q.M. for her collaboration, L. Pizzamiglio, G. Vallar and G. Antonucci for suggestions, C. Montagna, E. Catani and B. Fane for text revision, and G. Paganelli and his team for help in the radiological examination of Q.M. This work was supported by a grant from the Centro Residenziale IRCCS S. Lucia.

Correspondence and requests for materials should be addressed to F.D. at the IRCCS S. Lucia (e-mail: f.dorichi@uniroma1.it).

Glutamate mediates an inhibitory postsynaptic potential in dopamine neurons

C. D. Fiorillo & J. T. Williams

Vollum Institute, Oregon Health Sciences University, 3181 South West Sam Jackson Park Road, Portland, Oregon 97201, USA

Rapid information transfer within the brain depends on chemical signalling between neurons that is mediated primarily by glutamate and GABA (γ -aminobutyric acid), acting at ionotropic receptors to cause excitatory or inhibitory postsynaptic potentials (EPSPs or IPSPs), respectively. In addition, synaptically released glutamate acts on metabotropic receptors to excite neurons on a slower timescale through second-messenger cascades, including phosphoinositide hydrolysis¹. We now report a unique IPSP mediated by the activation of metabotropic glutamate receptors. In ventral midbrain dopamine neurons, activation of metabotropic glutamate receptors (mGluR1) mobilized calcium from caffeine/ryanodine-sensitive stores and increased an apamin-sensitive potassium conductance. The underlying potassium conductance and dependence on calcium stores set this IPSP apart from the slow IPSPs described so far^{2–4}. The mGluR-induced hyperpolarization was dependent on brief exposure to agonist, because prolonged application of exogenous agonist desensitized the hyperpolarization and caused the more commonly reported depolarization^{1,5,6}. The rapid rise and brief duration of synaptically released glutamate in the extracellular space can therefore mediate a rapid excitation through activation of ionotropic receptors, followed by inhibition through the mGluR1 receptor. Thus the idea that glutamate is solely an excitatory neurotransmitter must be replaced with a more complex view of its dual function in synaptic transmission.

Intracellular recordings were made in dopamine neurons of the ventral tegmental area (VTA) and substantia nigra. In the presence of ionotropic receptor antagonists, a train of stimuli (10 at 66 Hz) elicited a hyperpolarization that had two overlapping components of similar amplitude. The early component was mediated by GABA_B receptors (Fig. 1a), as previously reported⁷. The late component was insensitive to GABA_B antagonists (Fig. 1a, b) and reversibly blocked by tetrodotoxin (TTX, 300 nM; $91 \pm 2\%$, $n = 4$), suggesting that it might also be a slow IPSP.

The kinetics of this late hyperpolarization suggested that it was mediated by a mechanism distinct from GABA_B-mediated IPSP. The mGluR1-subtype is present on dopamine cells^{8,9}, is coupled to phosphoinositide (PtdIns) hydrolysis¹ and releases calcium from intracellular stores in other cells¹. Calcium could activate a calcium-sensitive potassium conductance (rSK3) present on dopamine cells that is sensitive to the bee venom toxin, apamin^{10,11}. Apamin irreversibly blocked the late hyperpolarization (by $86 \pm 3\%$, $n = 7$, Fig. 1a), but did not affect the GABA_B IPSP. Thus the late hyperpolarization was dependent on a rise in intracellular Ca^{2+} concentration and activation of a K^+ current.

The mGluR antagonists (S)-MCPG (0.5–1.0 mM), (S)-4-CPG (500 μM) and (RS)-AIDA (100–500 μM) reduced the amplitude of the late hyperpolarization by $75 \pm 5\%$ ($n = 6$, Fig. 1b, c), $73 \pm 7\%$ ($n = 4$), and $44 \pm 8\%$ ($n = 6$), respectively. This pharmacological profile suggests that the late hyperpolarization is an IPSP mediated by mGluR1^{12–14} and the mGluR1 splice variant has been specifically demonstrated in dopamine cells¹⁵. The inhibition caused was selective for the IPSP because (S)-MCPG and (S)-4-CPG did not affect the amplitude of the apamin-sensitive component of the after-hyperpolarization (AHP) following the action potential ((S)-

MCPG, $0 \pm 3\%$, $n = 5$; (S)-4-CPG, 11 ± 5 decrease, $n = 5$).

The mGluR-mediated IPSP had variable kinetics relative to the GABA_B IPSP (Fig. 1d). The peak latency ranged from 320 to 740 ms, with a mean value of 477 ± 13 ms ($n = 52$). IPSPs were frequently observed that had a long latency to onset (up to ~ 300 ms), but then rose to peak more rapidly than the GABA_B IPSP. The mGluR IPSP was also unique in that agents that altered the activation of mGluRs

also altered the peak latency. The mGluR1 antagonists MCPG and AIDA delayed the time to peak by 163 ± 38 ms ($n = 5$) and 48 ± 15 ms ($n = 6$), respectively. A μ -opioid (DAMGO, $1 \mu\text{M}$; $n = 6$) and a 5-HT (5-CT, 100 nM ; $n = 5$) agonist reduced the peak amplitude by $35 \pm 7\%$ and $23 \pm 4\%$, and delayed the peak latency by 24 ± 7 ms and 32 ± 11 ms, respectively. D1 receptor activation (SKF 82958, $1 \mu\text{M}$), which enhances glutamate release

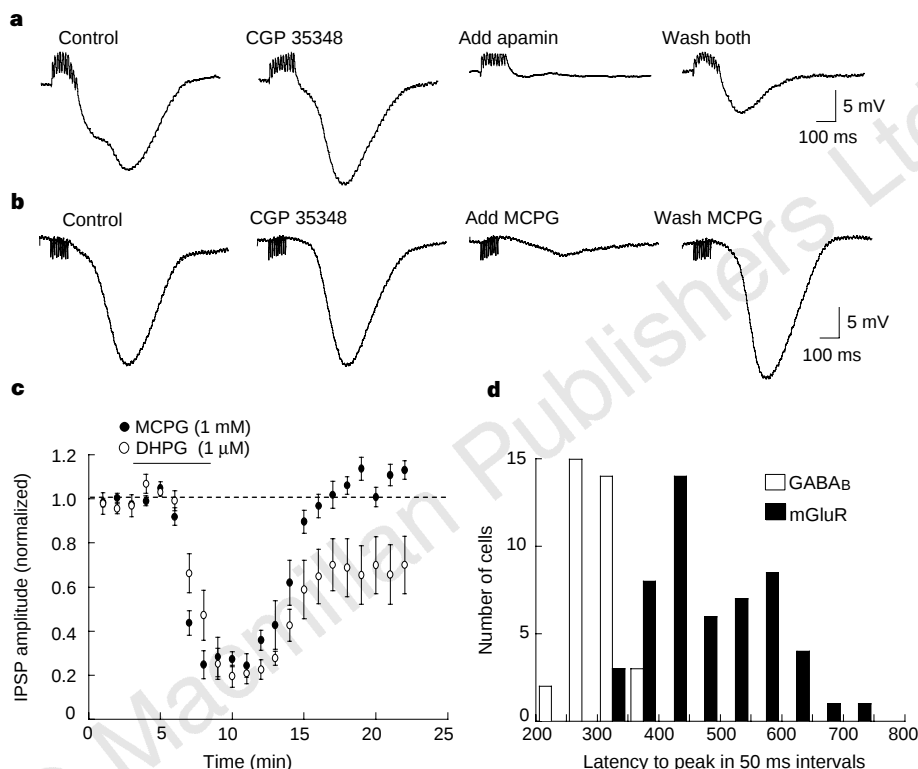


Figure 1 The late component of the IPSP is mediated by mGluRs and activation of an apamin-sensitive potassium conductance. **a**, Consecutive averaged traces showing that the early component of the IPSP is selectively blocked by the GABA_B antagonist CGP 35348 ($100 \mu\text{M}$), whereas the late component is selectively and irreversibly blocked by apamin (100 nM). Wash of apamin and CGP 35348 restores only the GABA_B IPSP. **b**, The late component of the IPSP is reversibly blocked by the mGluR antagonist (S)-MCPG (1 mM). **c**, The mGluR IPSP is blocked by (S)-MCPG ($0.5\text{--}1 \text{ mM}$, $n = 6$) and is desensitized by a low concentration of the mGluR

agonist (S)-DHPG ($1 \mu\text{M}$, $n = 5$). DHPG also caused a sustained depolarization of about 3 mV (see text). In all figures, each point represents the average of IPSPs from several cells, each of which was normalized to the average amplitude of the five IPSPs preceding drug application (MCPG, $-18.9 \pm 1.3 \text{ mV}$, $n = 6$; DHPG, $-17.8 \pm 3.1 \text{ mV}$, $n = 5$). **d**, Histogram of the peak latency for GABA_B ($-12.9 \pm 0.8 \text{ mV}$, $n = 34$) and mGluR ($-13.5 \pm 0.7 \text{ mV}$, $n = 52$) IPSPs. The peak latency was measured from the start of the stimulus train, which consisted of 10 stimuli and lasted 135 ms.

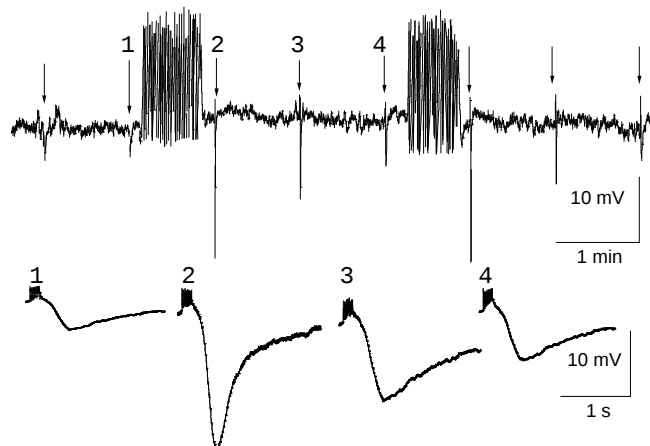


Figure 2 The amplitude of mGluR IPSPs is increased by prior depolarization of the membrane potential. Removing the holding current between stimuli allowed spontaneous action potentials and increased the amplitude of the IPSP. The full extent of the action potentials and the following after-hyperpolarizations are not

shown. Arrows indicate the points of stimulation. The downward deflections below the arrows represent IPSPs. The traces below show, on an expanded timescale, the IPSPs corresponding to the numbered arrows above.

in the VTA both *in vivo*¹⁶ and *in vitro* (P. Kalivas and D. Cameron, personal communication), augmented the IPSP ($32 \pm 5\%$, $n = 5$) and reduced the peak latency (46 ± 12 ms). None of these agonists had any effect on the membrane potential of dopamine cells. Agents that modulate the GABA_B IPSP in the same way do not alter its waveform^{17,18}.

While recording IPSPs, current was injected to prevent spontaneous action potentials, which normally allow calcium influx. In 11 of 25 neurons, the amplitude of the IPSP was augmented ($74 \pm 23\%$, 19–275%) and the peak latency reduced (79 ± 18 ms) by removing the holding current for 20 to 50 s between stimuli (Fig. 2). The facilitation often lasted for more than 1 min. Augmentation of the mGluR IPSP by prior depolarization may result from loading of intracellular Ca^{2+} stores that were partially depleted during prolonged hyperpolarization. We therefore tested agents known to disrupt the mobilization of Ca^{2+} from intracellular stores^{19,20}.

Ryanodine ($10 \mu\text{M}$), which blocks²¹ or reduces²² the conductance of the ryanodine receptor and prevents its activation, irreversibly reduced the mGluR IPSP by $75 \pm 12\%$ ($n = 6$; Fig. 3a) after 15–20 min, whereas it had no effect on the GABA_B IPSP ($4 \pm 8\%$ decrease, $n = 4$; Fig. 3a). Caffeine (10 mM), which activates the ryanodine receptor and may also interfere with activation of the inositol trisphosphate (InsP_3) receptor²⁰, rapidly and reversibly

inhibited the mGluR IPSP ($69 \pm 4\%$ after 4–6 min, $n = 9$; Fig. 3b), whereas it more slowly facilitated the GABA_B IPSP ($70 \pm 18\%$ after 6–10 min, $n = 4$; Fig. 3b). Thapsigargin²³ ($10 \mu\text{M}$) and cyclopiazonic acid²⁴ ($10 \mu\text{M}$), which deplete intracellular Ca^{2+} stores by blocking the ATPase that mediates Ca^{2+} uptake, blocked the mGluR IPSP by $19 \pm 5\%$ ($n = 5$) and $98 \pm 4\%$ ($n = 6$; Fig. 3c), respectively, after 15–20 min. The GABA_B IPSP was not altered by thapsigargin ($5 \pm 5\%$ increase, $n = 3$) or cyclopiazonic acid ($6 \pm 7\%$ decrease, $n = 4$; Fig. 3c).

Following the mGluR IPSP, a slow EPSP was occasionally observed. The slow EPSP had a longer peak latency (~ 800 ms) and duration, and was evoked with a greater number of stimuli than the mGluR IPSP (Fig. 4a). The slow EPSP was blocked by mGluR antagonists (MCPG or 4-CPG; $n = 4$), confirming reports of mGluR-mediated slow postsynaptic excitations in dopamine⁵ and other neurons¹. Thus, depending on the stimulation protocol, synaptic activation of mGluR1 can mediate both hyperpolarizing and depolarizing responses in the same cell.

Application of exogenous mGluR agonists was by superfusion or pressure ejection. Superfusion of the selective mGluR 1/5 agonists

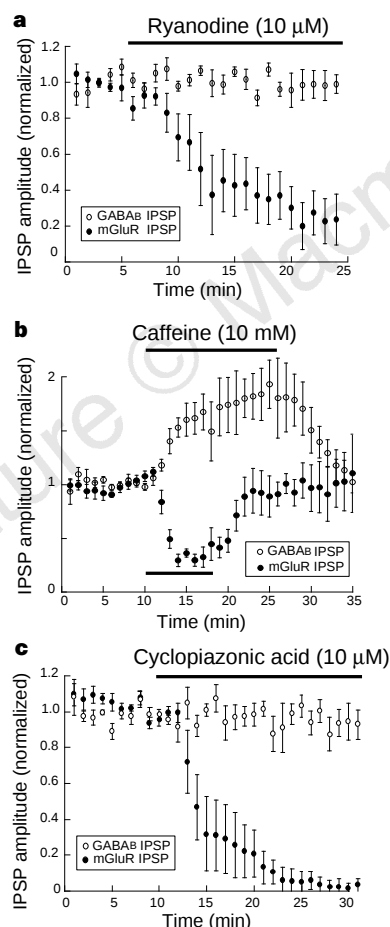


Figure 3 The mGluR-mediated IPSP is dependent on intracellular Ca^{2+} stores. **a**, Ryanodine ($10 \mu\text{M}$) blocked the mGluR IPSP (control = -12.9 ± 2.0 mV, $n = 6$) but had no effect on the GABA_B IPSP (control = -14.5 ± 2.3 mV, $n = 4$). **b**, Caffeine (10 mM) blocked the mGluR IPSP (control = -13.2 ± 1.9 mV, $n = 9$), whereas it facilitated the GABA_B IPSP (control = -11.2 ± 1.1 mV, $n = 4$). Caffeine was applied for a longer time on the GABA_B IPSP than on the mGluR IPSP, as indicated by the bars. **c**, Cyclopiazonic acid ($10 \mu\text{M}$) abolished the mGluR IPSP (control = -16.0 ± 1.8 , $n = 6$) without altering the GABA_B IPSP (control = -16.5 ± 3.7 , $n = 4$).

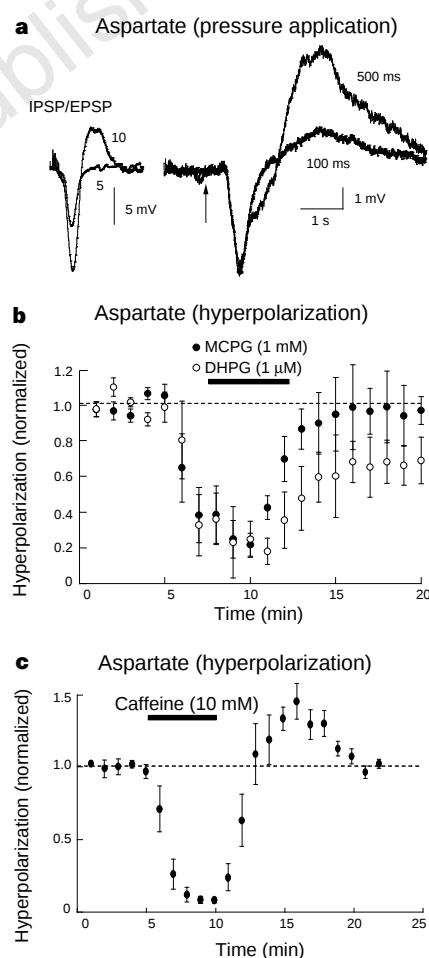


Figure 4 Brief exogenous application of aspartate applied by pressure ejection mimics the IPSP. **a**, Left, in this cell, a train of 5 stimuli evokes only an IPSP, whereas a train of 10 stimuli elicits an EPSP following the IPSP. Right, in another cell, pressure ejection of aspartate (1 mM) for 100 ms produced primarily a hyperpolarization, whereas a 500-ms ejection also produced a depolarization. The timescale is the same for traces on the left and right. **b**, The mGluR antagonist MCPG (1 mM) blocked and a low concentration of the agonist DHPG ($1 \mu\text{M}$) desensitized the hyperpolarizations produced by aspartate (1 mM) or glutamate ($100 \mu\text{M}$) (MCPG control = -9.0 ± 1.5 mV, $n = 4$; DHPG control = -8.7 ± 1.8 mV, $n = 4$). **c**, Caffeine (10 mM) also blocked the hyperpolarizations produced by aspartate (control = -9.4 ± 1.3 mV, $n = 4$). Shortly after washing out the caffeine, the response was transiently enhanced.

(S)-DHPG (1–50 μ M, $n = 16$) and quisqualate (2 μ M, $n = 5$) always depolarized the membrane, confirming previous studies showing that mGluR agonists depolarize dopamine⁶ and other neurons^{1,25}. One potential explanation for this apparent contradiction is that prolonged activation of mGluRs desensitizes the hyperpolarizing response. Agonists were therefore applied rapidly by pressure ejection to avoid desensitization.

Pressure ejections (5–800 ms) of L-glutamate (100 μ M) or L-aspartate^{25,26} (1 mM) caused both hyperpolarizations and depolarizations (Fig. 4a). Depolarizations were readily observed and blocked by mGluR antagonists, whereas hyperpolarizations were observed with more difficulty. In contrast to depolarizations, hyperpolarizations were only observed at lower concentrations of agonist and when applied closer to the soma. In all cells in which hyperpolarizations were observed ($n = 15$), they were elicited with shorter duration pulses than depolarizations (Fig. 4a). Hyperpolarizing responses were blocked by MCPG (1 mM, $80 \pm 9\%$, $n = 4$; Fig. 4b), caffeine (10 mM, $90 \pm 3\%$, $n = 4$; Fig. 4c) and apamin (100 nM, $92 \pm 1\%$, $n = 4$), but were resistant to TTX (300 mM, $n = 5$).

Desensitization of the hyperpolarization was examined with the combination of pressure ejection of aspartate and superfusion of DHPG (1 μ M) at a concentration expected to occupy only a fraction of mGluR1 receptors. DHPG (1 μ M) caused a sustained depolarization (3.4 ± 0.6 mV) and reduced the amplitude of the hyperpolarization induced by pressure ejection of aspartate by $80 \pm 11\%$ ($n = 4$; Fig. 4b), but did not affect the apamin-sensitive AHP ($8 \pm 6\%$ increase, $n = 6$). Superfusion of DHPG also caused an identical inhibition of the mGluR IPSP ($81 \pm 5\%$, $n = 5$; Fig. 1c). This suggests that continuous, low receptor occupancy rapidly desensitizes the hyperpolarizing response to mGluR activation downstream of the receptor itself.

Both in acutely dissociated hippocampal CA1 neurons and in cultured cerebellar granule cells, activation of mGluRs mobilizes intracellular calcium^{27,28} and activates a potassium conductance^{27,29}. In both cell types, the mGluR response was blocked by caffeine and ryanodine, and appeared to depend on activation of InsP₃ receptors^{27,28}. It is proposed that the mGluR-mediated IPSP results from activation of an mGluR1 that causes InsP₃-induced Ca²⁺ release, to trigger Ca²⁺-induced Ca²⁺ release through ryanodine receptors. The mGluR-mediated IPSP is therefore unlike previously characterized slow IPSPs, mediated by a membrane-delimited pathway and the activation of an inwardly rectifying potassium conductance^{2–5}. Not surprisingly, these two mechanisms give rise to IPSPs with distinct kinetic properties (Fig. 1d). It is unlikely that mGluRs are unique in evoking this form of synaptic inhibition, as we have observed similar hyperpolarizations mediated by PtdIns-coupled muscarinic acetylcholine receptors (unpublished results).

Like other PtdIns-coupled receptors, mGluR receptor activation has been generally found to be excitatory^{1,25}. Whereas most previous studies have used slow application of agonist, this study indicates that inhibition requires rapid application of agonist. Our results show that the activation of mGluR1 by synaptically released glutamate can induce Ca²⁺ mobilization in the absence of depolarization. In dopamine cells, this causes a pure inhibition by activation of a Ca²⁺-dependent potassium conductance. Prolonged synaptic activation of mGluRs may result in suppression of the IPSP as well as evoking a slow EPSP. Thus, depending on the frequency and pattern of afferent input to dopamine neurons, glutamate can mediate inhibition or excitation by activation of the same receptor. □

Methods

Slice preparation. Intracellular recordings were made in horizontal slices (300 μ m) of ventral midbrain from male Wistar rats (150–200 g). Details of the method of slice preparation and recording have been published¹⁸. Recordings were made from submerged slices in a chamber (0.5 ml) superfused with physiological saline at a rate of 1.5 ml min⁻¹ and maintained at 35°C. The solution was equilibrated with 95% O₂/5% CO₂ and contained (in mM): 126

NaCl, 2.5 KCl, 1.2 MgCl₂, 2.4 CaCl₂, 1.4 NaH₂PO₄, 25 NaHCO₃ and 11 D-glucose.

Recordings. Microelectrodes (50–80 M Ω) were filled with 2 M KCl. Recordings were made from dopamine neurons, identified by their electrical properties, in VTA (61) and substantia nigra (16). No differences were observed between neurons of the VTA and substantia nigra; the data were pooled. The membrane potential was held between –60 and –70 mV to prevent spontaneous action potentials.

Synaptic potentials. Synaptic potentials were evoked with bipolar tungsten stimulating electrodes with a tip separation of 300–1,000 μ m. A train of 10 stimuli of 400 μ s at 0.5–1.5 mA was delivered at 66 Hz every 60 s. Stimulating electrodes were placed within 1 mm rostral or caudal of the recording site.

All synaptic potentials were recorded in the presence of picrotoxin (100 μ M, GABA_A), strychnine (1 μ M, glycine), eticlopride (100 nM, D2), and CNQX (10 μ M, AMPA) or NBQX (5 μ M, AMPA). Slices were treated with MK-801 (50–100 μ M) at the start of all experiments to block NMDA-receptor-mediated currents. In all experiments on the mGluR IPSP, the solution also contained the GABA_B antagonist CGP 35348 (100–300 μ M), CGP 52432 (500–1,000 nM) or CGP 56999a (100–1,000 nM). Note that the GABA_A receptor antagonist bicuculline methiodide suppresses both the apamin-sensitive AHP³⁰ and the mGluR IPSP in the presence of picrotoxin.

Drugs. All drugs were applied by superfusion, except pressure ejections of L-glutamate (100 μ M) or L-aspartate (1 mM), which were performed at 20 p.s.i. with a Picospritzer II using glass pipettes (1 μ m tip diameter).

Data analysis. Values are given as arithmetic means $\pm$ s.e.m. All differences stated in the text are statistically significant (paired *t*-tests, $P < 0.05$). The amplitude of the after-hyperpolarization (AHP) following a single action potential was determined by measuring the difference in membrane potential between 20 ms before and 50 ms after the initiation of the action potential. Using this measurement, apamin blocked $63 \pm 4\%$ ($n = 8$) of the AHP, indicating the presence of one or more additional conductances underlying this potential change.

Received 8 January; accepted 20 April 1998.

- Pin, J.-P. & Duvoisin, R. The metabotropic glutamate receptors: structure and functions. *Neuropharmacology* **34**, 1–26 (1995).
- Otis, T. S., De Koninck, Y. & Mody, I. Characterization of synaptically elicited GABA_B responses using patch-clamp recordings in rat hippocampal slices. *J. Physiol.* **463**, 391–407 (1993).
- Surprenant, A. & Williams, J. T. Inhibitory synaptic potentials recorded from mammalian neurones prolonged by blockade of noradrenaline uptake. *J. Physiol.* **382**, 87–103 (1987).
- Pan, Z. Z., Colmers, W. F. & Williams, J. T. 5-HT-Mediated synaptic potentials in the dorsal raphe nucleus: interactions with excitatory amino acid and GABA neurotransmission. *J. Neurophysiol.* **62**, 481–486 (1989).
- Shen, K. Z. & Johnson, S. W. A slow excitatory postsynaptic current mediated by G-protein-coupled metabotropic glutamate receptors in rat ventral tegmental dopamine neurons. *Eur. J. Neurosci.* **9**, 48–54 (1997).
- Mercuri, N. B. et al. Activation of metabotropic glutamate receptors induces an inward current in rat dopamine mesencephalic neurons. *Neuroscience* **56**, 399–407 (1993).
- Johnson, S. W. & North, R. A. Two types of neurone in the rat ventral tegmental area and their synaptic inputs. *J. Physiol.* **450**, 455–468 (1992).
- Martin, L. J., Blackstone, C. D., Huganir, R. L. & Price, D. L. Cellular localization of a metabotropic glutamate receptor in rat brain. *Neuron* **9**, 259–270 (1992).
- Testa, C. M., Standaert, D. G., Young, A. B. & Penny, J. B. Jr Metabotropic glutamate receptor mRNA expression in the basal ganglia of the rat. *J. Neurosci.* **14**, 3005–3018 (1994).
- Kohler, M. et al. Small-conductance, calcium-activated potassium channels from mammalian brain. *Science* **273**, 1709–1714 (1996).
- Shepard, P. D. & Bunney, B. S. Repetitive firing properties of putative dopamine-containing neurons in vitro: regulation by an apamin-sensitive Ca²⁺-activated K⁺ conductance. *Exp. Brain Res.* **86**, 141–150 (1991).
- Brabet, L., Mary, S., Bockaert, J. & Pin, J.-P. Phenylglycine derivatives discriminate between mGluR1- and mGluR5-mediated responses. *Neuropharmacology* **34**, 895–903 (1995).
- Moroni, F. et al. Pharmacological characterization of 1-aminoindan-1,5-dicarboxylic acid, a potent mGluR1 antagonist. *J. Pharmacol. Exp. Ther.* **281**, 721–729 (1997).
- Conn, P. J. & Pin, J.-P. Pharmacology and functions of metabotropic glutamate receptors. *Annu. Rev. Pharmacol. Toxicol.* **37**, 205–237 (1997).
- Kolinski, C. M. et al. Expression of metabotropic glutamate receptor 1 (mGluR1) isoforms in the substantia nigra pars compacta of the rat. *Neuroscience* (in the press).
- Kalivas, P. W. & Duffy, P. D1 receptors modulate glutamate transmission in the ventral tegmental area. *J. Neurosci.* **15**, 5379–5388 (1995).
- Johnson, S. W., Mercuri, N. B. & North, R. A. 5-Hydroxytryptamine_{1B} receptors block the GABA_B synaptic potential in rat dopamine neurons. *J. Neurosci.* **12**, 2000–2006 (1992).
- Cameron, D. L. & Williams, J. T. Dopamine D1 receptors facilitate transmitter release. *Nature* **366**, 344–347 (1993).
- Kostyuk, P. & Verkhratsky, A. Calcium stores in neurons and glia. *Neuroscience* **63**, 381–404 (1994).
- Ehrlich, B. E., Kaftan, E., Bezprozvannaya, S. & Bezprozvannaya, I. The pharmacology of intracellular Ca²⁺ release channels. *Trends Pharmacol. Sci.* **15**, 145–149 (1994).
- McPherson, P. S. et al. The brain ryanodine receptor: a caffeine-sensitive calcium release channel. *Neuron* **7**, 17–25 (1991).
- Rousseau, E., Smith, J. S. & Meissner, G. Ryanodine modifies conductance and gating behavior of single Ca²⁺ release channel. *Am. J. Physiol.* **253**, C364–C368 (1987).

23. Thastrup, O. *et al.* Thapsigargin, a tumor promoter, discharges intracellular Ca^{2+} stores by specific inhibition of the endoplasmic reticulum Ca^{2+} -ATPase. *Proc. Natl. Acad. Sci. USA* **87**, 2466–2470 (1990).
24. Seidler, N. W., Jona, I., Vegh, M. & Martonosi, A. Cyclopiazonic acid is a specific inhibitor of the Ca^{2+} -ATPase of sarcoplasmic reticulum. *J. Biol. Chem.* **264**, 17816–17823 (1989).
25. Charpak, S., Gähwiler, B. H., Do, K. Q. & Knöpfel, T. Potassium conductances in hippocampal neurons blocked by excitatory amino-acid transmitters. *Nature* **347**, 765–767 (1990).
26. Masu, M. *et al.* Sequence and expression of a metabotropic glutamate receptor. *Nature* **349**, 760–765 (1991).
27. Shirasaki, T., Nobutoshi, H. & Akaike, N. Metabotropic glutamate response in acutely dissociated hippocampal CA1 pyramidal neurones of the rat. *J. Physiol.* **475**, 439–453 (1994).
28. Irving, A. J., Collingridge, G. L. & Schofield, J. G. Interactions between Ca^{2+} mobilizing mechanisms in cultured rat cerebellar granule cells. *J. Physiol.* **456**, 667–680 (1992).
29. Fagni, L., Bossu, J. L. & Bockaert, J. Activation of a large-conductance Ca^{2+} -dependent K^+ channel by stimulation of glutamate phosphoinositide-coupled receptors in cultured cerebellar granule cells. *Eur. J. Neurosci.* **3**, 778–789 (1991).
30. Johnson, S. W. & Seutin, V. Bicuculline methiodide potentiates NMDA-dependent burst firing in rat dopamine neurons by blocking apamin-sensitive Ca^{2+} -activated K^+ currents. *Neurosci. Lett.* **231**, 13–16 (1997).

Acknowledgements. We thank O. J. Manzoni, P. Chavis, G. L. Westbrook and N. V. Marrion for comments on the manuscript. This work was supported by the National Institute on Drug Abuse.

Correspondence and requests for materials should be addressed to J.T.W. (e-mail: williamj@ohsu.edu).

Wingless and Notch regulate cell-cycle arrest in the developing *Drosophila* wing

Laura A. Johnston & Bruce A. Edgar

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA

In developing organs, the regulation of cell proliferation and patterning of cell fates is coordinated. How this coordination is achieved, however, is unknown. In the developing *Drosophila* wing, both cell proliferation and patterning require the secreted morphogen Wingless (Wg) at the dorsoventral compartment boundary (reviewed in ref. 1). Late in wing development, Wg also induces a zone of non-proliferating cells at the dorsoventral boundary. This zone gives rise to sensory bristles of the adult wing margin^{2,3}. Here we investigate how Wg coordinates the cell cycle with patterning by studying the regulation of this growth arrest. We show that Wg, in conjunction with Notch, induces arrest in both the G1 and G2 phases of the cell cycle in separate subdomains of the zone of non-proliferating cells. Wg induces G2 arrest in two subdomains by inducing the proneural genes *achaete* and *scute*, which downregulate the mitosis-inducing phosphatase String (*Cdc25*)⁴. Notch activity creates a third domain by preventing arrest at G2 in *wg*-expressing cells, resulting in their arrest in G1.

The zone of non-proliferating cells (ZNC) is composed of three subdomains, each about four cells wide. Cells in the central domain express *wg* (Fig. 1a)⁵. This central domain is flanked by dorsal and ventral domains which, in the anterior compartment, express the proneural transcription factors *Achaete* (Ac) and *Scute* (Sc) (Fig. 2d)^{6,7}. Cells of the ZNC stop proliferating 30 h before most other cells in the disc but re-enter the cell cycle for two or three divisions after pupariation^{2,8}. The arrest is seen by an absence of cells in S phase or mitosis (Fig. 1b, f). The domain architecture of the ZNC is suggested by the expression patterns of *string* (*stg*) and of the G2 cyclins A (*CycA*) and B (*CycB*)⁹. In the anterior compartment, cells in the dorsal and ventral domains do not express *stg* messenger RNA but accumulate high levels of G2 cyclins in the cytoplasm (Fig. 1d, e). As *Stg* is required for mitosis⁴ and *Stg* and the G2 cyclins are degraded at cell division⁹, these patterns are indicative of arrest in G2. In contrast, in the central domain *CycA* and *CycB* proteins are undetectable (Fig. 1d), but *stg* mRNA is expressed (Fig. 1e). This indicates that these central cells may be arrested in G1. To confirm the phases in which cells of the different subdomains are arrested, we induced expression of the mitotic inducer *Stg* or of the S-phase inducer cyclin E (*CycE*)¹⁰. Within an hour after inducing *stg*

expression, cells in the anterior dorsal and ventral domains entered mitosis (Fig. 1g). This indicates that these two domains become arrested in G2 because of an absence of *stg*. Induction of *cycE* led to entry into S phase only in the central domain in the anterior ZNC, but in the posterior ZNC most cells entered S phase (Fig. 1h). We conclude that the anterior central domain and the entire posterior ZNC are arrested in G1, whereas the dorsal and ventral domains of the anterior compartment are arrested in G2 (Fig. 1j).

Several observations indicate that the G1 arrest may be due to inactivation of dE2F, a factor required to activate the transcription of genes needed for DNA replication¹¹. First, although dE2F is expressed throughout the ZNC (data not shown), its activity appears to be downregulated because one of its targets, the ribonucleotide reductase-2 gene (*RNR-2*)¹¹, is not expressed there (Fig. 1c). Second, forced expression of *dE2F* bypassed the G1 arrest and also induced expression of *RNR-2* in the ZNC (data not shown). *CycE* is also normally present throughout the ZNC (data not shown) but does not promote S-phase progression, indicating that repression of *CycE* activity may also have a role in the G1 arrest.

How does Wg activity regulate the G1 and G2 arrests in the ZNC? Wg activates transcription of target genes through its signal-transducing proteins, Armadillo/ β -catenin (Arm) and dTCF/Lef-1 (reviewed in ref. 12). We eliminated Wg activity either by using a temperature-sensitive *wg* allele or by expressing a dominant-negative form of dTCF throughout the ZNC. Loss of *wg* function during the last 24–30 h of disc development abolished both the G1 and G2 arrests (Fig. 2a) and allowed *stg* expression in the anterior dorsal and ventral domains (Fig. 2b). To determine how Wg normally represses

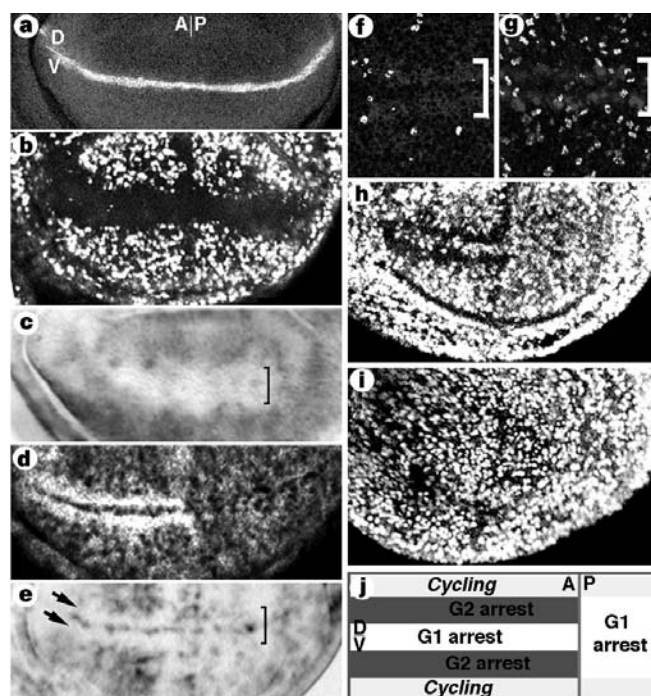


Figure 1 The ZNC contains cells arrested in G1 or G2 phase. **a**, Anti-Wg staining. D/V, dorsoventral boundary; A/P, anteroposterior boundary. All discs are orientated similarly. **b**, The ZNC does not incorporate BrdU. **c**, *RNR-2* mRNA is absent from the ZNC. **d**, *CycA* accumulates to high levels in the two anterior domains. *CycB* accumulates similarly. **e**, *stg* mRNA is absent (arrows) from anterior dorsal and ventral cells of the ZNC, but present in posterior and anterior central cells. **f**, Wild-type ZNC, stained for mitotic cells. **g**, *stg* overproduction induces mitoses in the dorsal and ventral anterior domains. The ZNC is marked by brackets, mitotic cells are white, and the dorsal and ventral domains are light grey in **f**. **g**, **h**, *cycE* overproduction induces BrdU incorporation in all posterior cells of the ZNC, but only in the central domain in the anterior compartment. **i**, BrdU labelling showing that simultaneous induction of *stg* and *cycE* bypasses both G1 and G2 arrests. **j**, Summary of cell-cycle arrests in the ZNC.

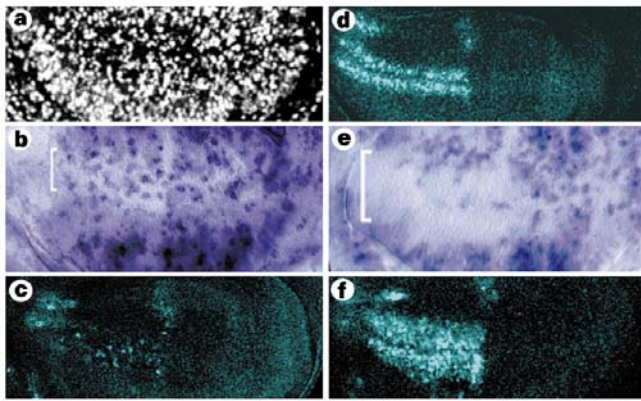


Figure 2 Wg regulates *stg* in the G2-arrested cells. Sections are orientated as in Fig. 1. **a**, Ectopic expression of dominant-negative dTCF (dTCF^{ΔN}) causes loss of arrest and BrdU incorporation. Use of the temperature-sensitive *wg* allele at the restrictive temperature produced identical results. dTCF^{ΔN} results in *stg* expression where it is normally repressed (**b**, bracket) and causes loss of *ac* expression (**c**). **d**, Wild-type *ac* expression. **e**, Ectopic Wg expression in all three domains of the ZNC widens the domain of *stg* repression (bracket). **f**, *ac* expression is widened dorsally and ventrally by overproduction of Wg.

stg in these domains, we broadened the effects of *wg* by expressing Wg or a constitutively active form of Arm in all three domains of the ZNC. Both treatments caused more cells to lose *stg* expression and arrest in G2, including many central domain cells which usually express *stg* (Fig. 2e). Ectopic Wg expression also expanded the G2 arrest dorsally and ventrally, indicating that the dorsal and ventral limits of the ZNC reflect a specific concentration threshold of Wg signalling.

Four observations suggested that the proneural genes *ac* and *sc* regulated the G2 arrest of the ZNC. First, *ac* and *sc* are expressed in the *stg*-negative, G2-arrested cells, but not in G1-arrested cells (Fig. 2d). Second, loss of Wg activity results in loss of expression of *ac* and *sc* (Fig. 2c)³ and causes expression of *stg* (Fig. 2b). Third, ectopic Wg or activated Arm expands the domain of *stg*-negative, Ac-positive cells (Fig. 2e, f). Fourth, expression of *ac* and *sc* in the ZNC is extinguished just before re-entry into the cell cycle after pupariation^{7,13}. We therefore examined *stg* expression in the ZNC of discs from animals lacking both *ac* and *sc* function. Although the ZNC of these discs did not incorporate bromodeoxyuridine (BrdU) (Fig. 3a)¹⁴, many cells expressed *stg* mRNA, indicating that the ZNC was probably exclusively G1-arrested (data not shown). This was

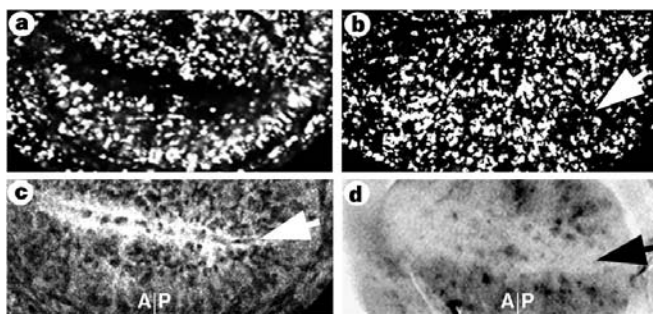


Figure 3 Ac and Sc repress *stg* transcription. Sections are orientated as in Fig. 1. **a**, BrdU labelling of *sc*^{10.1} discs shows a ZNC. **b**, Overproduction of CycE in *sc*^{10.1} discs allows BrdU incorporation throughout the ZNC (arrow), indicating that this ZNC consisted previously of only G1-arrested cells. **c**, **d**, Overproduction of Scute arrests both anterior and posterior cells of the ZNC in G2. **c**, Scute causes CycA to accumulate in both anterior and posterior cells (arrow; compare with Fig. 1d). **d**, *stg* mRNA is repressed in both anterior and posterior cells of the ZNC after overproduction of Scute (arrow; compare with Fig. 1e).

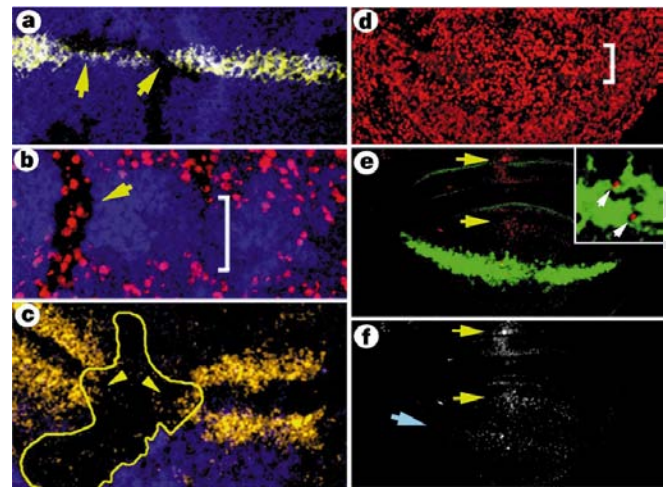


Figure 4 Notch activity prevents G2 arrest. **a**, Disc stained for Wg protein (yellow); Wg is not expressed in clones lacking *Su(H)* activity (arrows). **b**, BrdU incorporation (red) in *Su(H)*-deficient clones (lacking blue staining). **c**, CycA expression (yellow) in a *Su(H)*-deficient clone (inside the outline) in the ZNC. Only *Su(H)*-deficient cells adjacent to Wg-expressing cells (outside the clone) accumulate CycA (arrowheads). **d**, Overproduction of *cycE* in discs expressing activated Notch (N^{ΔE}) leads to BrdU incorporation (red) throughout the ZNC. **e**, Anti-Ac staining, showing expression outside the ZNC (red; arrows), but not where N^{ΔE} is expressed (green). Insert, magnification of two Ac-positive cells in the ZNC, showing no overlap with N^{ΔE} expression. **f**, Single-channel image of disc in **e** showing lack of *ac* expression in the ZNC (blue arrow; compare with Fig. 2d).

confirmed by overproduction of *cycE* in the *ac*- and *sc*-null discs; this led to S-phase entry throughout the ZNC (Fig. 3b). Thus, in the absence of *ac* and *sc* function, all cells in the ZNC arrested in G1. Conversely, forced expression of Sc throughout the ZNC resulted in loss of *stg* expression and cytoplasmic accumulation of CycA, indicating G2 arrest, in both posterior and anterior compartments (Fig. 3c, d). Together, these results indicate that Wg controls the G2 arrest in the ZNC by inducing expression of *ac* and *sc*, which in turn downregulate *stg*.

Notch is highly activated in the *wg*-expressing cells of the ZNC¹⁵ and is required for *wg* expression there^{16,17}. To determine whether Notch also regulates the cell-cycle arrests in the ZNC, we generated clones of cells that lacked the activity of *Suppressor of Hairless* (*Su(H)*), a transcription factor that transduces the Notch signal¹⁸. When such clones were located in the central domain of the ZNC, they lost expression of *wg* (Fig. 4a)¹⁹ and many cells entered S phase

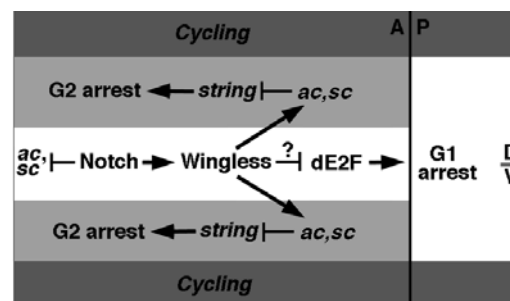


Figure 5 Model of cell-cycle arrests in the ZNC. Wingless induces a G2 arrest in dorsal and ventral anterior cells by inducing *ac* and *sc*, which repress *string* expression (light grey domains). At the same time Wingless inactivates dE2F throughout the ZNC, leading to G1 arrest in the posterior and anterior central domains (white domains). Notch promotes both G1 and G2 arrests by sustaining Wingless expression, and also creates the anterior domain architecture by blocking *ac* and *sc* expression in Wingless-expressing cells. A, anterior; P, posterior; D, dorsal; V, ventral.

(Fig. 4b). However, when cells within anterior clones were located adjacent to wild-type cells that expressed Wg, they accumulated CycA (Fig. 4c) and did not incorporate BrdU, indicating G2 arrest. Cells lacking *Su(H)* activity located a distance from *wg*-expressing cells did not accumulate CycA, and neither did posterior clones. Thus, cells lacking the ability to transduce Notch signalling either entered S phase or, if anterior and adjacent to Wg, they arrested in G2.

To test whether G1 arrest requires Notch activity, we expressed a constitutively active form of Notch throughout the ZNC. In these discs the ZNC was still present. However, *stg* was expressed and CycA did not accumulate in anterior cells, indicating that although the cells were not cycling, activated Notch prevented the G2 arrest (not shown). Overproduction of *cycE* in discs expressing activated Notch led to S phase throughout the ZNC (compare Figs 4d, 1h), confirming that cells in the ZNC had been entirely G1-arrested when activated Notch was expressed alone. In addition, activated Notch severely reduced expression of Ac in the ZNC (Fig. 4e, f). As *ac* and *sc* are required for the G2 arrest, we conclude that expression of activated Notch prevents G2 arrest by blocking expression of Ac and, presumably, Sc. These results indicate that expression of Ac and Sc, and G2 arrest, in the central domain is normally prevented by the high level of endogenous Notch activity there. To test whether Notch directly regulates the G1 arrest, we constructed discs lacking Wg activity, but expressing activated Notch in the ZNC. These discs did not form a ZNC at all (data not shown). Thus, in the absence of Wg activity, Notch is not sufficient to induce a G1 arrest.

Our observations lead to a model of how signalling by Wg controls a bipartite cell-cycle arrest in the wing disc (Fig. 5). Late in disc development, Wg protein emanating from the dorsoventral boundary inhibits dE2F activity throughout the ZNC region, promoting a G1 arrest. In the anterior dorsal and ventral domains, Wg also induces expression of Ac and Sc, leading to repression of *stg* transcription and therefore to G2 arrest in these cells. The *stg* promoter region contains putative Ac/Sc-binding sites, indicating that these basic helix-loop-helix proteins could repress *stg* transcription directly (L.A.J. and B.A.E., unpublished data). Our results indicate that Notch controls neither the G1 nor G2 arrest directly, but that it generates diversity within the ZNC in two ways. First, Notch indirectly promotes both arrests by maintaining *wg* expression^{16,19}. Second, Notch blocks the G2 arrest in the central domain by inhibiting *ac/sc* expression. This function of Notch is also essential to inhibit the acquisition of neural fates²⁰. Thus, one mechanism that patterning signals such as Wg use to control cell proliferation is the induction of target genes, for example *ac* and *sc*, that regulate both the cell cycle and cell-fate specification. Intriguingly, the dorsoventral-boundary region of the developing vertebrate spinal cord also contains a ZNC, and the corresponding region in *Xenopus* expresses the *ac/sc* homologue *Xash-3* (C. Queva and K. Sharma, personal communications)²¹. Thus, the insect and vertebrate cell-cycle arrests may be produced by similar mechanisms. The presence of a constriction at the ZNC in the spinal cord indicates that zones of non-proliferating cells may act to constrain tissue growth and provides a mechanism for morphogenesis. □

Methods

Mutant stocks. We removed both Ac and Sc functions using the *sc*^{10.1} allele²². In some experiments Wg activity was reduced using the temperature-sensitive allele *wg*^{IL114} (ref. 23) at 25°C. *wg*^{cx4} is a null allele of *wg*²⁴.

Immunocytochemistry. BrdU labelling and immunocytochemistry were performed as described²⁵. A protocol for *in situ* hybridization is available on request.

Transgenes and construction of recombinant strains. We used the Gal4/UAS (upstream activation sequence) system²⁶ to express transgenes specifically in the ZNC. Gal4^{C96} (ref. 27) is expressed in the wing disc from approximately 80 h of development, before ZNC formation, and is expressed in cells in all three domains. The following UAS transgenes were expressed: dominant-

negative dTCF²⁸, activated Arm²⁹, Wg (gift from J.-P. Vincent), Scute (gift from C. Doe), CycE (gift from C. Lehner), dE2F, Stg, or activated Notch (N^{ΔE}; gift from E. Giniger). We also used heat-shock transgenes (HS) of CycE, Stg, or dE2F. Larvae were heat-shocked at 37°C for 1 h, and discs were dissected following a 1–2 h recovery period. To eliminate Wg function in discs expressing activated Notch, we constructed *w; wg^{cx4}/UASdTCF^{ΔN}*; Gal4^{C96}/UAS N^{ΔE} animals. To express CycE in animals with activated Notch, HS-CycE was recombined onto the N^{ΔE} chromosome to make *w; UAS-N^{ΔE}, HS-CycE*/Gal4^{C96} animals. CycE was overproduced in the ZNC of *sc*^{10.1} animals with *sc*^{10.1}/Y; Gal4^{C96}/UAS-CycE.

Mitotic recombination. Clones of cells lacking *Su(H)* activity were generated by mitotic recombination induced with the Flp/FRT system³⁰ in HSFlp; FRT^{40A} *arm*lacZ (or HS-*π*-myc)/FRT^{40A} *Su(H)*^{SF8} larvae during the second instar.

Received 16 December 1997; accepted 23 April 1998.

- Serrano, N. & O'Farrell, P. H. Limb morphogenesis: connections between patterning and growth. *Curr. Biol.* **7**, 186–195 (1997).
- O'Brochta, D. A. & Bryant, P. J. A zone of non-proliferating cells at a lineage restriction boundary in *Drosophila*. *Nature* **313**, 138–141 (1985).
- Phillips, R. G. & Whittle, J. R. *wingless* expression mediates determination of peripheral nervous system elements in late stages of *Drosophila* wing disc development. *Development* **118**, 427–438 (1993).
- Edgar, B. A. & O'Farrell, P. H. The three postblastoderm cell cycles of *Drosophila* embryogenesis are regulated in G2 by *string*. *Cell* **62**, 469–480 (1990).
- Blair, S. S. Mechanisms of compartment formation: evidence that non-proliferating cells do not play a critical role in defining the D/V lineage restriction in the developing wing of *Drosophila*. *Development* **119**, 339–351 (1993).
- Cubas, P., de Celis, J. F., Campuzano, S. & Modolell, J. Proneural clusters of *achaete-scute* expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. *Gene Dev.* **5**, 996–1008 (1991).
- Skeath, J. B. & Carroll, S. B. Regulation of *achaete-scute* gene expression and sensory organ pattern formation in the *Drosophila* wing. *Genes Dev.* **5**, 984–995 (1991).
- Hartenstein, V. & Posakony, J. W. Development of adult sensilla on the wing and notum of *Drosophila melanogaster*. *Development* **107**, 389–405 (1989).
- Lehner, C. F. & O'Farrell, P. H. The roles of *Drosophila* cyclins A and B in mitotic control. *Cell* **61**, 535–547 (1990).
- Knoblich, J. A. *et al.* Cyclin E controls S phase progression and its down-regulation during *Drosophila* embryogenesis is required for the arrest of cell proliferation. *Cell* **77**, 107–120 (1994).
- Duronio, R. J., O'Farrell, P. H., Xie, J. E., Brook, A. & Dyson, N. The transcription factor E2F is required for S phase during *Drosophila* embryogenesis. *Genes Dev.* **9**, 1445–1455 (1995).
- Nusse, R. A versatile transcriptional effector of *Wingless* signaling. *Cell* **89**, 321–323 (1997).
- Romani, S., Campuzano, S., Macagno, E. R. & Modolell, J. Expression of *achaete* and *scute* genes in *Drosophila* imaginal discs and their function in sensory organ development. *Sensory Dev.* **3**, 997–1007 (1989).
- Usui, K. & Kimura, K.-I. Sensory mother cells are selected from among mitotically quiescent clusters of cells in the wing disc of *Drosophila*. *Development* **116**, 601–610 (1992).
- de Celis, J. F., Garcia-Bellido, A. & Bray, S. J. Activation and function of Notch at the dorsal-ventral boundary of the wing imaginal disc. *Development* **122**, 359–369 (1996).
- Couso, J. P., Knust, E. & Martinez, A. A. Serrate and *Wingless* cooperate to induce *vestigial* gene expression and wing formation in *Drosophila*. *Curr. Biol.* **5**, 1437–1448 (1995).
- Rulifson, E. J. & Blair, S. S. Notch regulates *wingless* expression and is not required for reception of the paracrine *wingless* signal during wing margin neurogenesis in *Drosophila*. *Development* **121**, 2813–2824 (1995).
- Schweisguth, F. Suppressor of Hairless is required for signal reception during lateral inhibition in the *Drosophila* pupal notum. *Development* **121**, 1875–1884 (1995).
- Neumann, C. J. & Cohen, S. M. A hierarchy of cross-regulation involving *Notch*, *wingless*, *vestigial* and *cut* organizes the dorsal/ventral axis of the *Drosophila* wing. *Development* **122**, 3477–3485 (1996).
- Kopan, R. & Turner, D. L. The Notch pathway: democracy and aristocracy in the selection of cell fate. *Curr. Opin. Neurobiol.* **6**, 594–601 (1996).
- Zimmerman, K., Shih, J., Bars, J., Collazo, A. & Anderson, D. J. *Xash-3*, a novel *Xenopus* *achaete-scute* homolog, provides an early marker of planar neural induction and position along the mediolateral axis of the neural plate. *Development* **119**, 221–232 (1993).
- Campuzano, S. *et al.* Molecular genetics of the *achaete-scute* complex of *D. melanogaster*. *Cell* **40**, 327–338 (1985).
- Bejsovec, A. & Martinez, A. A. Roles of *wingless* in patterning the larval epidermis of *Drosophila*. *Development* **113**, 471–485 (1991).
- Baker, N. E. Molecular cloning of sequences from *wingless*, a segment polarity gene in *Drosophila*: the spatial distribution of a transcript in embryos. *EMBO J.* **6**, 1765–1773 (1987).
- Johnston, L. A. & Schubiger, G. Ectopic expression of *wingless* in imaginal discs interferes with *decapentaplegic* expression and alters cell determination. *Development* **122**, 3519–3529 (1996).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Gustafson, K. & Boulianne, G. L. Distinct expression patterns detected within individual tissues by the Gal4 enhancer trap technique. *Genome* **39**, 174–182 (1996).
- van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *DTCF*. *Cell* **88**, 789–799 (1997).
- Pai, L. M. *et al.* *Drosophila* alpha-catenin and E-cadherin bind to distinct regions of *Drosophila* Armadillo. *J. Cell Biol.* **271**, 32411–32420 (1996).
- Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).

Acknowledgements. We thank T. Neufeld, A. Bejsovec, M. Peifer, J.-P. Vincent, C. Doe, S. Blair, S. Carroll, R. Nusse, S. Parkhurst, H. Richardson, C. Lehner, G. Boulianne and E. Giniger for gifts of flies or antibodies, C. Queva and K. Sharma for communicating unpublished results, and R. Kopan, M. Schubiger, S. Blair, J. Overbaugh, J. Priess, M. Emerman, J. Roberts and members of the Edgar lab for discussions and comments on the manuscript. L.A.J. is supported by a grant from the NIH. B.A.E. receives support from the NIH, and is a Rita Allen and a Lucille P. Markey Scholar.

Correspondence and requests for materials should be addressed to L.A.J. (e-mail: lajohnst@thrcr.org).

Tout-velu is a *Drosophila* homologue of the putative tumour suppressor *EXT-1* and is needed for Hh diffusion

Yohanns Bellaïche*, Inge The*† & Norbert Perrimon*†

* Department of Genetics and † Howard Hughes Medical Institute, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

Hedgehog (Hh) proteins act through both short-range and long-range signalling to pattern tissues during invertebrate and vertebrate development¹. The mechanisms allowing Hedgehog to diffuse over a long distance and to exert its long-range effects are not understood. Here we identify a new *Drosophila* gene, named *tout-velu*, that is required for diffusion of Hedgehog. Characterization of *tout-velu* shows that it encodes an integral membrane protein that belongs to the *EXT* gene family. Members of this family are involved in the human multiple exostoses syndrome, which affects bone morphogenesis^{2–4}. Our results, together with the previous characterization of the role of

Indian Hedgehog in bone morphogenesis^{5–7}, lead us to propose that the multiple exostoses syndrome is associated with abnormal diffusion of Hedgehog proteins. These results show the existence of a new conserved mechanism required for diffusion of Hedgehog.

We identified mutations in a new gene, *tout-velu* (*ttv*), which means 'all hair', in a screen for maternal-effect mutations associated with *Drosophila* segment polarity⁸. The mutant segment-polarity phenotype is reminiscent of the phenotypes of *wingless* (*wg*) or *hedgehog* (*hh*) mutant embryos. As several specific *wg*-dependent processes are not affected in embryos that lack both maternal and zygotic *ttv* gene activities, it is likely that *ttv* is required for Hh, but not for *Wg*, signalling (I.T. *et al.*, manuscript in preparation). We analysed the function of *ttv* in the wing imaginal disc to determine its possible role in the Hh signalling pathway.

In the wing disc, Hh is expressed in cells of the posterior compartment and diffuses into the anterior compartment^{9–12}. In the anterior compartment, Hh signalling results in expression of Patched (*Ptc*)¹³, which can be detected as far as five cells away from the anterior/posterior boundary, and in stabilization of Cubitus interruptus (*Ci*), which is observed as far as 8–10 cells away (Fig 1a, b; refs. 14, 15). We therefore used *Ptc* expression and *Ci* stabilization as reporters with which to examine the role of *ttv* in Hh signalling. We generated somatic mutant clones with the *ttv*^[(2)00681] mutation, which is either a null allele or a severe hypomorphic allele (see Methods; I.T. *et al.*, manuscript in preparation). In large anterior *ttv*

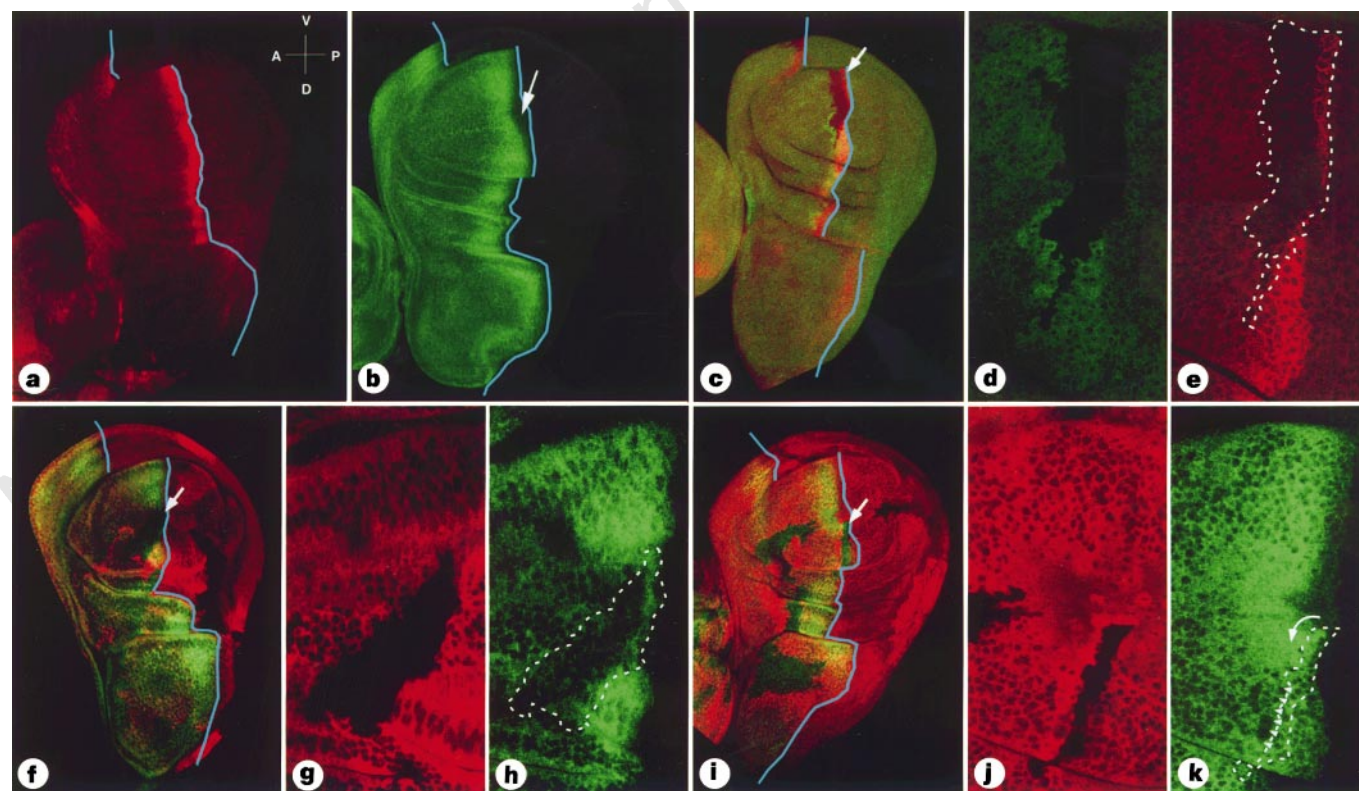


Figure 1 *ttv* affects Hh signalling. All imaginal discs are orientated as in **a**. A, anterior; P, posterior; V, ventral; D, dorsal. Blue lines indicate the A/P boundary. Dashed lines indicate the limit of the relevant clones as determined by the absence of β-galactosidase staining (see Methods). **a**, Wild-type *Ptc* expression (red)^{25,26}. **b**, Wild-type *Ci* expression (green) is detected with antibody 2A1 which has a higher affinity for the full-length *Ci* gene product¹⁵. *Ci* is only expressed in the anterior compartment. The posterior limit of the domain of *Ci* stabilization does not correspond with the anterior/posterior boundary (see arrow)¹⁴. **c–i**, *ttv* mutant clones. **c, f, i** Full discs; the relevant *ttv* mutant clones are indicated by an arrow. **d, e, g, h, j, k**, Magnifications of the relevant clones. **c–e**, *Ptc* expression in red (**c, e**) and β-galactosidase expression in green (**c, d**). **d, e**, In a *ttv* clone, *Ptc*

expression is only weakly induced, in one row of cells adjacent to the anterior/posterior boundary. **f–k**, *Ci* staining (green) (**f, h, i, k**) and β-Gal staining (red) (**f, g, i, j**). **g, h**, *Ci* stabilization is induced only at the posterior edge of the clone. **j, k**, Wild-type cells anterior to a *ttv* clone do not stabilize *Ci*, although they are in a domain that is competent to respond to Hh (arrowheads); thus, the non-cell-autonomous effect is directional (see also **h**). Within the clone, the distance between the domain of *Ci* stabilization and the A/P boundary is reduced, indicating that the level of Hh signalling is reduced¹⁴. We interpret the stabilization of *Ci* in the ventral cells of the *ttv* clone as being due to the diffusion of Hh from ventral wild-type cells (see arrow).

mutant clones adjacent to the anterior/posterior border, Hh signalling is impaired, as indicated by the lack of Ptc staining in most of the clone (Fig. 1c, e). However, Ptc expression still occurs at the posterior edge of the clone next to wild-type cells (Fig. 1e). The level of signalling in these cells is nevertheless diminished, as shown by a

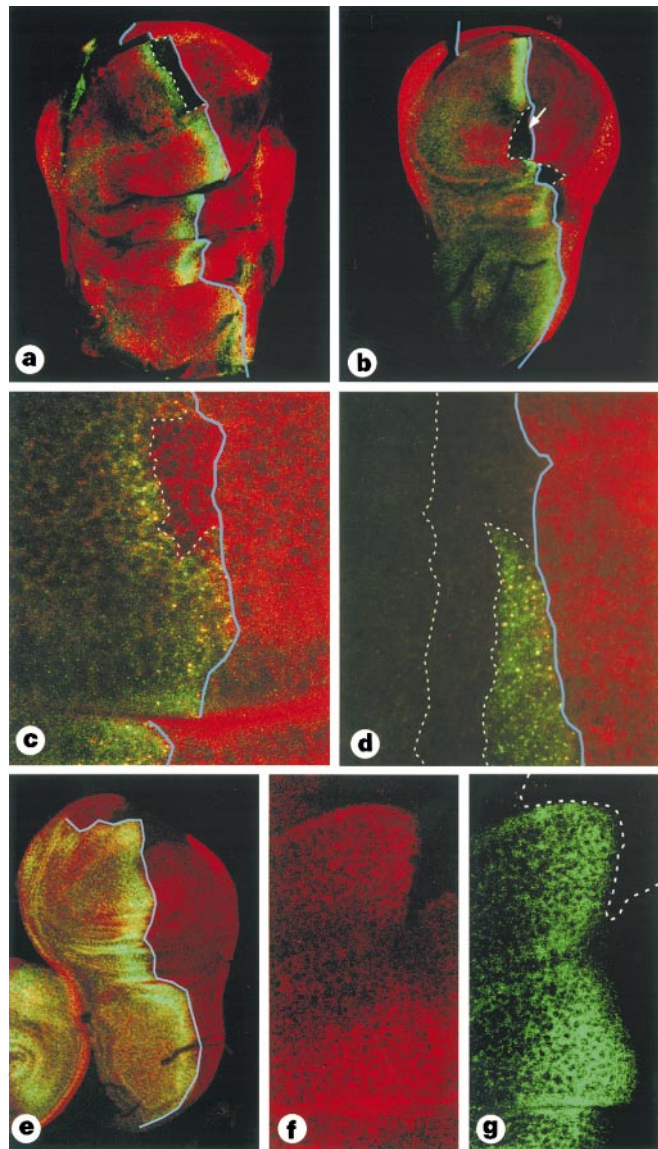


Figure 2 *ttv* is required for Hh diffusion. Blue lines indicate anterior/posterior boundary. Dashed lines indicate clone limits. Orientation is as in Fig. 1a. **a, b**, Overlays of Ptc (green) and β -galactosidase (red) stainings. **a**, A *ptc* mutant clone at the anterior/posterior boundary causes ectopic expression of Ptc in wild-type cells anterior to the clone (see also ref. 12). **b**, Two independent *ptc ttv* mutant clones in the anterior (see arrow) and posterior compartments. In an anterior *ptc ttv* clone, *ptc* expression is not induced anterior to the clone. In a posterior *ptc ttv* clone, Hh diffusion is not affected as *ptc* expression is induced. **c, d**, Overlays of Ptc (green) and Hh (red) staining²⁷. **c**, The *ptc* clone is in the anterior compartment. Hh staining is detected in, and anterior to, the *ptc* clone. **d**, Hh staining cannot be detected in a *ptc ttv* clone located in the anterior compartment. Clone boundaries are identified by β -galactosidase staining (not shown). **e–g**, Ci (green) (**e, g**) and β -galactosidase (red) staining (**e, f**). **e**, Full disc with a posterior clone. **f, g**, Magnification of the clone. Loss of *ttv* activity in the posterior compartment does not affect the diffusion of Hh in the anterior compartment, as Ci is stabilized. We consistently see a reduction of the domain of Ci stabilization. Hh signalling probably depends on the cumulative expression of Hh from several rows of cells adjacent to the anterior/posterior boundary. In a posterior *ttv* clone, only Hh produced by the row of cells adjacent to the anterior/posterior boundary can diffuse into the anterior compartment and contribute to signalling.

reduction in their levels of Ptc staining compared with wild-type cells (Fig. 1e).

A similar result was observed when Ci levels were used as a reporter for Hh signalling (Fig. 1f, h). In addition, the large region of Ci stabilization allowed us to determine a non-cell-autonomous effect of *ttv* mutant clones. When *ttv* mutant clones of only a few cells wide are located along the anterior/posterior boundary, wild-type cells anterior to the *ttv* mutant clone do not respond to the Hh signal, as shown by Ci staining (Fig. 1i, k). As these cells are located within a domain to which Hh is normally able to diffuse, and in which Hh can normally induce the stabilization of Ci, we conclude that *ttv* mutant clones have a directional, non-cell-autonomous effect on wild-type cells located anteriorly. As the domain of Ci stabilization is under direct control of Hh signalling^{14,15}, we interpret this result as an inability of Hh to reach wild-type cells located anteriorly to *ttv* mutant clones. We then investigated whether diffusion of Hh is impaired in the absence of *ttv*.

In a *ptc* mutant clone, Hh diffusion is observed as an ectopic induction of Ptc in wild-type cells localized anteriorly to a clone of *ptc* mutant cells (Fig. 2a; ref. 12). To determine whether Hh diffusion is modified by *ttv* in the *ptc* mutant clones, we generated clones with mutations in both *ptc* and *ttv* and analysed the expression of Ptc. In these double mutant clones, Ptc expression is not induced in wild-type cells anterior to the clone (Fig. 2b). To assess Hh diffusion in the *ptc ttv* mutant clones directly, we compared the distribution of Hh in *ptc* and *ptc ttv* clones. In *ptc* mutant cells, we detected Hh protein as diffuse membrane staining. When Hh reaches wild-type cells beyond a *ptc* clone, it can be seen in a punctate staining pattern that, for the most part, coincides with the punctate Ptc staining that reflects Ptc localization in vesicles (Fig. 2c). In *ptc ttv* double mutant clones, there is no Hh staining in mutant cells (Fig. 2d). On the basis of this result and the directional non-cell-autonomous effect of *ttv* mutant clones, we propose that Hh cannot diffuse in the absence of *ttv* activity.

In order to diffuse, Hh must move from the sending to the receiving cell; therefore, *ttv* could function in the sending cell and/or the receiving cell. In the anterior cell, the sending of Hh depends on its reception from the previous cells; thus it is difficult to determine where Ttv functions by analysing *ttv* mutant clones in the anterior compartment. We therefore generated mutant clones in the posterior compartment next to the anterior/posterior boundary and analysed their effects on Ci stabilization. We found that *ttv* mutant clones in the posterior compartment do not affect the diffusion of Hh (Fig. 2b, e, g). Thus, we propose that *ttv* functions in the receiving cells for the movement of Hh from sending to receiving cells. This observation also indicates that *ttv* is not required for Hh production. In a *ttv* clone located in the anterior compartment, Hh can still signal, although less efficiently, to the first receiving cells. We suggest that this weakened signalling activity is mediated by Hh present on the membrane of the Hh-sending cells and that efficient Hh signalling, even to adjacent cells, requires diffusion of Hh.

To clone the *ttv* gene, we isolated the genomic region surrounding the inserted P-element, *ttv*⁽²⁾⁰⁰⁶⁸¹ (Fig. 3a), and identified two genes in the vicinity of the P-element insertion site. One gene is *lamin C*, which is unlikely to correspond to *ttv* as it is not expressed maternally¹⁶. Sequencing of the putative 3.8-kilobase (kb) complementary DNA encoding *ttv* shows that Ttv is a protein of 760 amino acids that is similar to the human multiple exostoses (EXT-1) protein² (Fig. 3b). The 3.8-kb cDNA identifies a maternally expressed transcript on an RNA blot. A transgene of this transcript under the control of a heat-shock protein 70 gene (*hsp70*) promoter rescues *ttv* homozygous flies to viability (data not shown).

Ttv is 56% and 25% identical to the human EXT-1 and EXT-2 proteins, respectively. EXT-1 and EXT-2 probably form a new family of conserved molecules, as mouse and *Caenorhabditis elegans* homologues have been identified¹⁷. However, the cellular

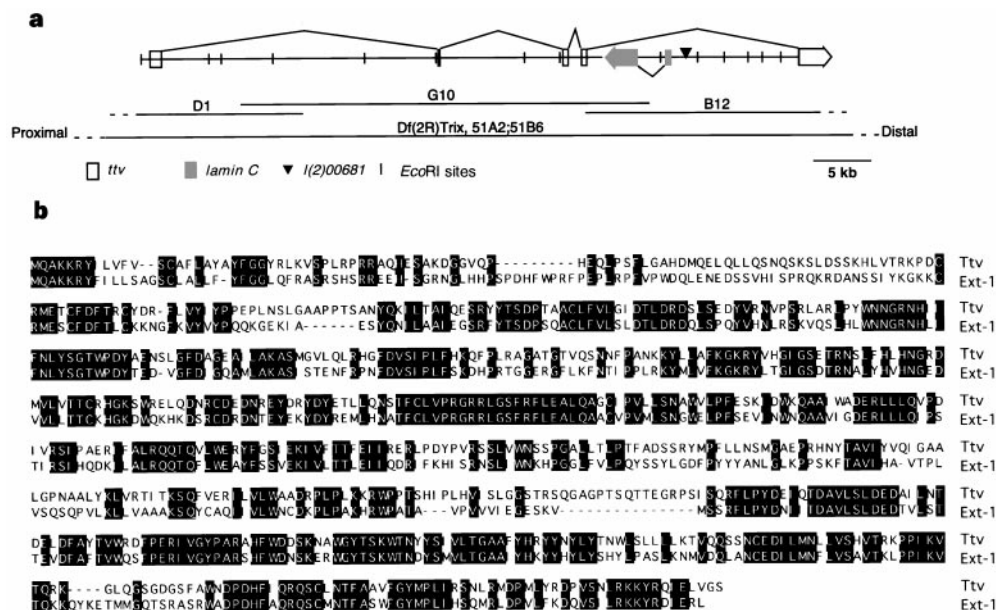


Figure 3 Cloning of the *ttv* locus. **a**, Genomic map of the *ttv* locus. The P-element, *ttv*^{I(2)00681} (black triangle), maps to the 51A polytene chromosome band which is deleted in *Df(2R)Trix*. The P-element is inserted into the sixth intron of the *ttv* gene, which also includes the *lamin C* gene. Only introns of more than 1 kb in length are

shown. D1, G10 and B12 are three overlapping cosmids that were previously mapped to the region²⁸. **b**, Sequence comparison between the Ttv and human EXT-1 proteins. Black boxes indicate amino acids that are identical in EXT-1 and Ttv. The overall identity between Ext-1 and Ttv is 56%.

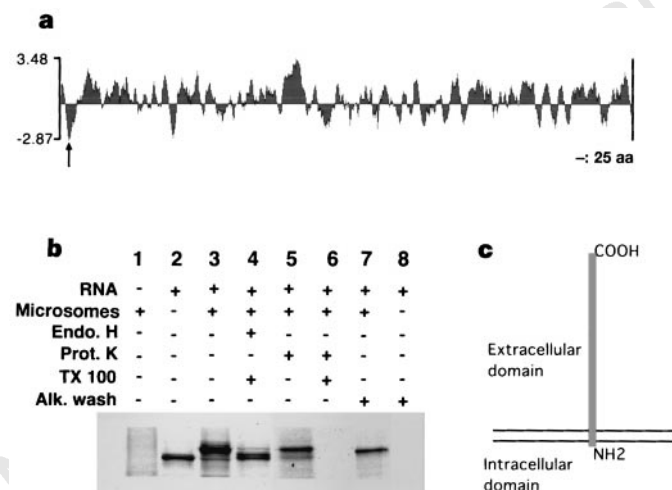


Figure 4 Ttv is a type II integral membrane protein. **a**, The Ttv Kyte-Doolittle hydrophobicity plot shows a hydrophobic stretch at the N terminus of the protein (amino acids (aa) 7-24, see arrow). **b**, *In vitro* translation and microsome assays. *In vitro* transcribed *ttv* cDNA is translated in the presence of rabbit reticulocyte extract (lane 2). When the same extract is supplemented with dog pancreatic microsomes, a mobility shift is detected (lane 3). This shift can be abolished by treatment with endoglycosidase H (lane 4). The full-length Ttv protein enter microsomes: Ttv is protected from digestion by proteinase K in the presence of intact microsomes (lanes 5, 6). Ttv remains associated with the membrane fraction after alkaline wash (lanes 7, 8), indicating that Ttv is a membrane protein. **c**, Proposed structure of the Ttv protein. The cytoplasmic tail is only six amino acids in length.

compartment in which EXT molecules could function has remained unknown^{2,18}. Analysis of the Ttv sequence shows that there is a hydrophobic stretch at the amino terminus of the Ttv protein, indicating that Ttv might be a transmembrane protein (Fig. 4a). We used an *in vitro* translation assay to show that Ttv is indeed a membrane protein (Fig. 4b). *In vitro*-transcribed *ttv* messenger RNA is translated into a putative protein with a relative molecular

mass (M_r) of ~80K. In the presence of microsomes, the protein is glycosylated and fully protected from degradation by proteinase K, indicating that the protein is imported into microsomes. Furthermore, the Ttv protein remains associated with the membrane fraction after an alkaline wash; it is therefore a membrane-associated protein¹⁹. As the sequence surrounding the signal sequence does not closely match consensus cleavage sequences²⁰, we propose that the signal sequence is not cleaved and instead acts as an anchor region. We conclude that Ttv is a type II integral membrane protein (Fig. 4c).

As the active form of Hh is tethered to the membrane by a cholesterol moiety^{21,22}, it is not known how Hh can diffuse from one cell to another. The movement of Hh does not depend on Ptc and Smoothened¹². As Ttv is needed in the receiving cell to allow Hh diffusion, we suggest that Ttv does not function in the dissociation of Hh from the membrane of the sending cells and is probably to be required to allow reassociation or maintenance of Hh on the surface of the receiving cells. The multiple exostoses syndrome is a dominantly inherited disease that is characterized by short stature, limb length inequalities, bone deformities and the presence of bone outgrowths, called exostoses, at the ends of long bones⁴. The loss of any of the *EXT-1*, -2 or -3 loci is responsible for this syndrome⁴. The *EXT-1* and *EXT-2* genes have been cloned and may be tumour-suppressor genes²⁻⁴. During bone morphogenesis one of the established roles of Indian Hh is to limit the rate of chondrocyte differentiation⁵⁻⁷. On the basis of the similarity between *EXT-1* and Ttv, we propose that some aspects of the multiple exostoses syndrome are due to defects in diffusion and efficient signalling of Indian Hh. The characterization of Ttv is an important step toward the understanding of a cellular mechanism that allows Hh to diffuse and to exert its patterning activity. □

Methods

Somatic clonal analyses. The *ttv*^{I(2)00681} allele behaves as a genetic null allele: the phenotype of embryos derived from *ttv*^{I(2)00681} germline clones (GLCs)⁸ crossed with *ttv*^{I(2)00681}/Cyo GLCs is identical to the phenotype of embryos derived from *ttv*^{I(2)00681} GLCs crossed with *Df(2R)Trix*/Cyo GLCs. Somatic clones were induced using the FRT/FLP-mediated recombination system²³. Mutant clones were identified by the loss of β -galactosidase staining, except

in Fig. 2c where the clone was identified by loss of Ptc staining. Relevant genotypes: for *ttv* clones: *hsFLP; FRT^{G13}ttv⁽²⁾⁰⁰⁶⁸¹/FRT^{G13} arm-LacZ*, for *ptc* clones: *hsFLP; FRT^{42D}ptc^{11W}/FRT^{42D} arm-lacZ* (*ptc^{11W}* is a protein-null allele²⁴); and for *ptc ttv* double mutant clones: *hsFLP; FRT^{42D}ptc^{11W}ttv⁽²⁾⁰⁰⁶⁸¹/FRT^{42D} arm-LacZ* (ref. 1). First-instar larvae were heat-shocked for 1 h at 37 °C. Third-instar larval imaginal discs fixed for 20 min in 4% formaldehyde in PBS with 0.1% Tween20 (PBT). Antibodies were diluted in PBT: rabbit anti-β-galactosidase (Cappel), 1:4,000 dilution; rat anti-β-galactosidase, 1:100; mouse anti-Ptc, 1:100; rabbit anti-Hh, 1:2,000 (TSA amplification was used to enhance the signal; NEN); rat anti-Ci, 1:5. Secondary antibodies coupled to fluorescent isothiocyanate, Texas Red or Cy5 (Jackson) were used at 1:200 dilution. Secondary antibodies coupled to horseradish peroxidase for TSA amplification (Vector) were used at 1:2,000 dilution. Stainings were visualized using a Leica TCS/NT confocal microscope.

Rescue construct. A *SpeI*, *NotI* fragment of the *ttv* gene was cloned downstream of the *hsp70* promoter of the pCasper-hsp vector and used to transform the *y* w strain using the helper plasmid pp25.1. One insertion on the third chromosome rescued *ttv* homozygous mutants to viability without heat shock. *ttv⁽²⁾⁰⁰⁶⁸¹* and two other alleles, *ttv⁽²⁾⁰⁰⁶⁶¹⁰⁹* and *ttv⁽²⁾⁰²⁰⁵⁵*, were tested.

In vitro translation assay. *ttv* RNA was produced by *in vitro* transcription using the SP6 RNA machine kit (Ambion); rabbit reticulocyte lysates, microsomes and endoglycosidase H were purchased from Boehringer. *In vitro* translation assays in the presence of 10 μCi ³⁵S-labelled methionine, proteinase K digestion (0.1 mg ml⁻¹ per reaction) and endoglycosidase H treatments (2 μU per reaction) were performed according to manufacturers' protocols. The samples were analysed by SDS-PAGE. The alkaline wash was performed as described¹⁹.

Sequence analysis. The hydrophobic plot was generated using Protean software. Analysis of the peptide cleavage site was done on the PSORT II web server (<http://psort.nibb.ac.jp:80/form2.html>) using an algorithm derived from ref. 20.

Received 6 February; accepted 16 April 1998.

- Johnson, R. L. & Tabin, C. The long and short of Hedgehog signaling. *Cell* **81**, 313–316 (1995).
- Ahn, J. et al. Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nature Genet.* **11**, 137–143 (1995).
- Stickens, D. & Evans, G. A. Isolation and characterization of the murine homolog of the human EXT2 multiple exostoses gene. *Biochem. Mol. Med.* **61**, 16–21 (1997).
- Wicklund, C. L., Pauli, R. M., Johnston, D. & Hech, J. T. Natural history study of hereditary multiple exostoses. *Am. J. Med. Genet.* **55**, 43–46 (1995).
- Lanske, B. et al. PTH/PTHrP receptor in early development and Indian Hedgehog-regulated bone growth. *Science* **273**, 663–666 (1996).
- Vortkamp, A. et al. Regulation of rate of cartilage differentiation by Indian Hedgehog and PTH-related protein. *Science* **273**, 613–622 (1996).
- Vortkamp, A. et al. Recapitulation of signals regulating embryonic bone formation during postnatal growth and in fracture repair. *Mech. Dev.* **71**, 65–76 (1998).
- Perrimon, N., Lanjuin, A., Arnold, C. & Noll, E. Zygotic lethal mutations with maternal effect phenotypes in *Drosophila melanogaster*. II. Loci on the second and third chromosomes identified by P-element-induced mutations. *Genetics* **144**, 1681–1692 (1996).
- Lee, J. J., von Kessler, D. P., Parks, S. & Beachy, P. A. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell* **71**, 33–50 (1992).
- Tabata, T. & Kornberg, T. B. Hedgehog is a signaling protein with a key role in patterning *Drosophila* imaginal discs. *Cell* **76**, 89–102 (1994).
- Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by Hedgehog protein. *Nature* **368**, 208–214 (1994).
- Chen, Y. & Struhl, G. Dual roles for patched in sequestering and transducing Hedgehog. *Cell* **87**, 553–563 (1996).
- Alexandre, C., Jacinto, A. & Ingham, P. W. Transcriptional activation of Hedgehog target genes in *Drosophila* is mediated directly by the Cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes Dev.* **10**, 2003–2013 (1996).
- Strigini, M. & Cohen, S. M. A Hedgehog activity gradient contributes to AP axial patterning of the *Drosophila* wing. *Development* **124**, 4697–4705 (1997).
- Aza-Blanc, P., Ramirez-Weber, F. A., Laget, M. P., Schwartz, C. & Kornberg, T. B. Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043–1053 (1997).
- Riemer, D. et al. Expression of *Drosophila* lamin C is developmentally regulated: analogies with vertebrate A-type lamins. *J. Cell Sci.* **108**, 3189–3198 (1995).
- Clines, G. A., Ashley, J. A., Shah, S. & Lovett, M. The structure of the human multiple exostoses 2 gene and characterization of homologs in mouse and *Caenorhabditis elegans*. *Genome Res.* **7**, 359–367 (1997).
- Stickens, D. et al. The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. *Nature Genet.* **14**, 25–32 (1996).
- Fujiki, Y., Hubbard, A. L., Fowler, S. & Lazarow, P. B. Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J. Cell Biol.* **93**, 97–102 (1982).
- von Heijne, G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683–90 (1986).
- Porter, J. A., Young, K. E. & Beachy, P. A. Cholesterol modification of Hedgehog signaling proteins in animal development. *Science* **274**, 255–259 (1996).
- Porter, J. A. et al. Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoproteolytic domain. *Cell* **86**, 21–34 (1996).
- Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1995).

- Nakano, Y. et al. A protein with several possible membrane-spanning domains encoded by the *Drosophila* segment polarity gene *patched*. *Nature* **341**, 508–513 (1989).
- Mullor, J. L., Calleja, M., Capdevila, J. & Guerrero, I. Hedgehog activity, independent of decapentaplegic, participates in wing disc patterning. *Development* **124**, 1227–1237 (1997).
- Capdevila, J., Pariente, E., Sampedro, J., Alonso, J. & Guerrero, I. Subcellular localization of the segment polarity protein Patched suggests an interaction with the wingless reception complex in *Drosophila* embryos. *Development* **120** (suppl), 987–998 (1994).
- Taylor, A. M., Nakano, Y., Mohler, J. & Ingham, P. W. Contrasting distributions of patched and Hedgehog proteins in the *Drosophila* embryo. *Mech. Dev.* **42**, 89–96 (1993).
- Dickson, B. J., Dominguez, M., van der Straten, A. & Hafen, E. Control of *Drosophila* photoreceptor cell fates by phyloped, a novel nuclear protein acting downstream of the Raf kinase. *Cell* **80**, 453–462 (1995).

Acknowledgements. We thank I. Guerrero, E. Hafen, B. Holmgren, P. Ingham, W. Mothes, G. Struhl and D. Raden for strains or reagents; A. Manoukian for his initial contribution to this work; W. Mothes for help with the microscope assays; B. Seed, C. Tabin and S. Goode for discussions; and S. Goode, B. Matthey-Prevot, S. van den Heuvel, M. Zeidler, E. Bach and M. Pettit for comments on the manuscript. Y.B. thanks A. Morineau for encouragement during this study. Y.B. is supported by the Boehringer Ingelheim Fonds; I.T. is supported by the Damon Runyon Fellowship. This work was supported by the NSF and the HHMI where N.P. is an investigator.

Correspondence and requests for materials should be addressed to N.P. (e-mail: perrimon@rascal.med.harvard.edu).

A dimeric 14-3-3 protein is an essential cofactor for Raf kinase activity

Guri Tzivion, Zhijun Luo & Joseph Avruch

Diabetes Unit and Medical Services and the Department of Molecular Biology, Massachusetts General Hospital, and the Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA

cRaf-1 is a mitogen-activated protein kinase that is the main effector recruited by GTP-bound Ras in order to activate the MAP kinase pathway¹. Inactive Raf is found in the cytosol in a complex with Hsp90, Hsp50 (Cdc37)^{2,3} and the 14-3-3 proteins⁴. GTP-bound Ras binds Raf and is necessary but not sufficient for the stable activation of Raf that occurs in response to serum, epidermal growth factor, platelet-derived growth factor or insulin^{5–8}. These agents cause a two- to threefold increase in overall phosphorylation of Raf on serine/threonine residues^{8,9}, and treatment of cRaf-1 with protein (serine/threonine) phosphatases can deactivate it, at least partially¹⁰. The role of 14-3-3 proteins in the regulation of Raf's kinase activity is uncertain^{4,11} and is investigated here. Active Raf can be almost completely deactivated *in vitro* by displacement of 14-3-3 using synthetic phosphopeptides. Deactivation can be substantially reversed by addition of purified recombinant bacterial 14-3-3; however, Raf must have been previously activated *in vivo* to be reactivated by 14-3-3 *in vitro*. The ability of 14-3-3 to support Raf activity is dependent on phosphorylation of serine residues on Raf and on the integrity of the 14-3-3 dimer; mutant monomeric forms of 14-3-3, although able to bind Raf *in vivo*, do not enable Raf to be activated *in vivo* or restore Raf activity after displacement of 14-3-3 *in vitro*. The 14-3-3 protein is not required to induce dimerization of Raf. We propose that dimeric 14-3-3 is needed both to maintain Raf in an inactive state in the absence of GTP-bound Ras and to stabilize an active conformation of Raf produced during activation *in vivo*.

Table 1 Interaction of Raf variants with 14-3-3, Ras and MEK-1 in a yeast two-hybrid assay

	Colour produced in an assay with		
	14-3-3†	MEK-1	Ras
Raf (WT)	Blue	Blue	Blue
Raf (S259A)	Blue	Blue	Blue
Raf (S259A/S621A)	White	Blue	Blue

A blue colour in the two-hybrid assay indicates that a protein-protein interaction occurs.

We have attempted to clarify the role of 14-3-3 in regulating Raf activity by studying both the effect of displacing 14-3-3 from Raf using synthetic phosphopeptides and the behaviour of Raf mutants that have an altered ability to bind to 14-3-3. Synthetic phosphopeptides based on amino-acid sequences surrounding major sites of phosphorylation of Raf (such as Ser 621) displaced Raf from 14-3-3 (Fig. 1a). Displacement of 14-3-3 from Raf was accompanied by deactivation of the kinase activity of Raf (Fig. 1b); both basal and EGF-stimulated activity of glutathione-S-transferase (GST)-Raf

Phosphopeptide also inhibited highly purified baculoviral recombinant Raf, which was shown previously to purify in a complex with 14-3-3, Hsp90 and Hsp50 (ref. 18). Phosphopeptide inhibited kinase activity in a dose-dependent manner whether Raf was activated *in vivo* by expression with the V12Ras mutant, v-Src, or both V12 Ras and v-Src (Fig. 1d). The unphosphorylated peptide caused little or no inhibition and the presence of phosphatase inhibitors did not alter significantly the ability of the phosphopeptide to inhibit baculoviral recombinant Raf (Fig. 1d, right); even the addition of a highly active recombinant protein phosphatase-1 did not diminish Raf activity in the conditions used (Fig. 2c: compare lanes 2, 9).

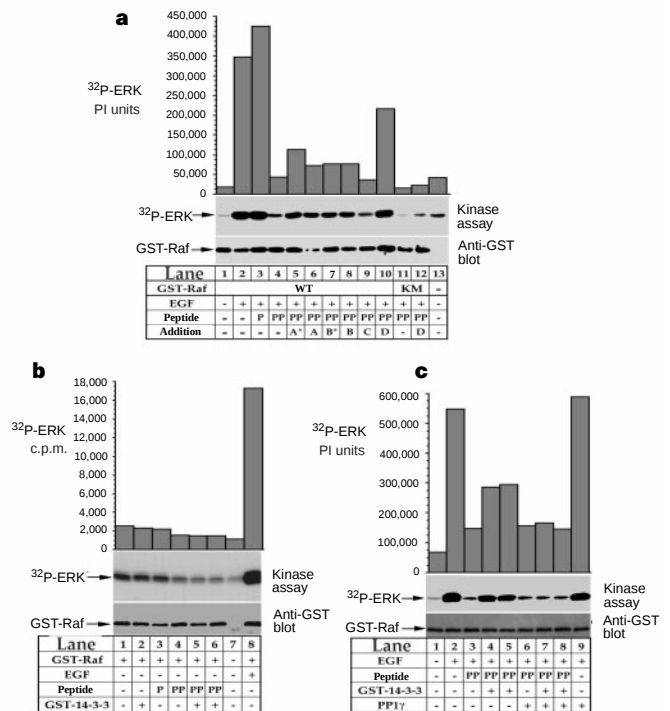


Figure 2 Raf inactivated by phosphopeptide *in vitro* can be reactivated by recombinant 14-3-3. **a**, Extracts of EGF-stimulated COS cells expressing GST-Raf (WT) or an inactive mutant GST-Raf (Lys375 → Met; KM) were incubated with vehicle or peptides (P or PP) as in Fig. 1a. Purified, immobilized, phosphopeptide-treated GST-Raf or KM was incubated for 90 min with extracts of COS cells, which were treated with carrier or EGF (asterisk indicates EGF treatment) and transiently expressed recombinant Myc/14-3-3 (A,A*) no Myc/14-3-3 (B,B*), purified recombinant GST (15 μg ml⁻¹; C) or purified recombinant GST/14-3-3 (15 μg ml⁻¹; D). Reaction mixtures were then washed and assayed for kinase activity. **b**, Extracts from serum-deprived COS cells expressing GST-Raf were incubated with vehicle or peptides (P or PP) as in Fig. 1a. After purification, immobilized GST-Raf polypeptides were incubated with buffer or bacterial recombinant GST/14-3-3 (15 μg ml⁻¹) for 90 min, washed and assayed for Raf kinase activity. GST-Raf activity after addition of EGF is shown in lane 8. **c**, GST-Raf extracted from EGF-treated COS cells was incubated with vehicle or phosphopeptide as in Fig. 1a. After purification, immobilized GST-Raf was incubated for 60 min with buffer or protein phosphatase-1 (PP1γ, 1 mU ml⁻¹), washed and further incubated with carrier or bacterial recombinant GST/14-3-3 as in **a**. This was followed with assay of Raf kinase activity. The activity of GST-Raf from serum-deprived cells is shown in lane 1. PI units. Phospho-Imager units.

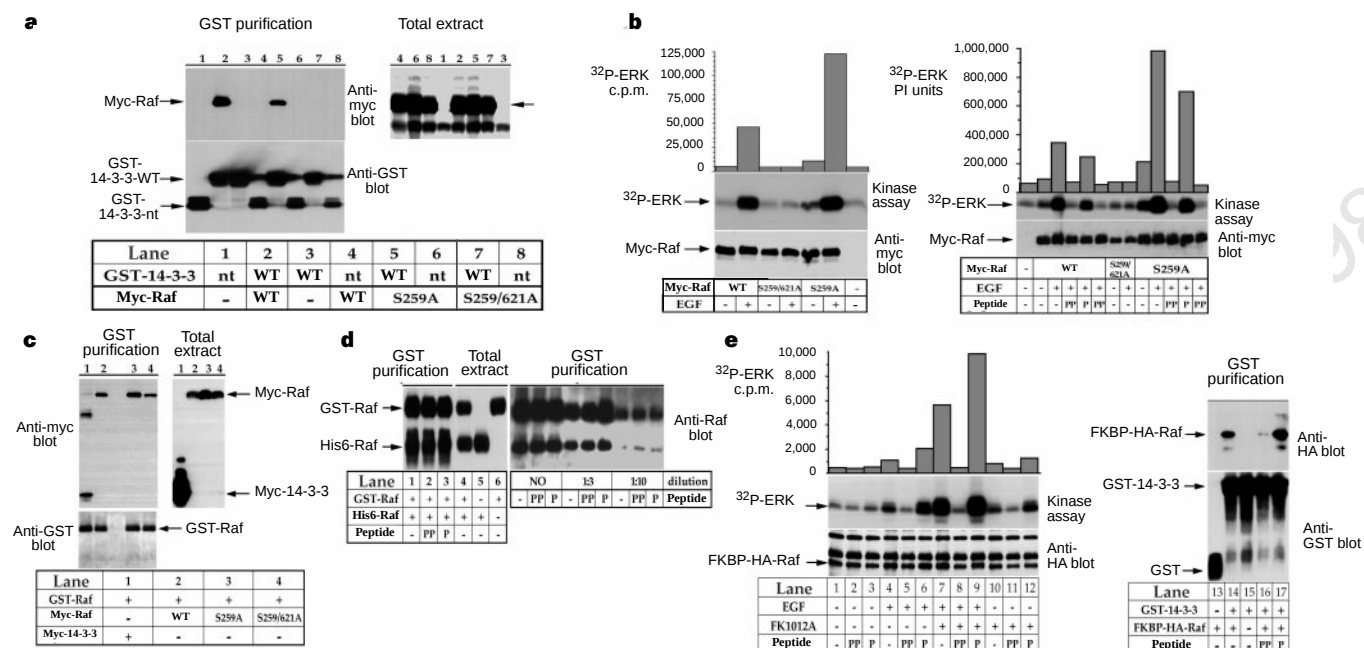


Figure 3 The binding of 14-3-3 to Raf is necessary for Raf activity but not for Raf oligomerization. **a**, GST/14-3-3 (WT) or GST/14-3-3(1-140)(nt) was expressed in COS cells with Myc-cRaf-1 (WT), Myc-Raf(S259A) or Myc-Raf(S259A/S621A). The purified GST-fusion proteins (left) and the extracts (right) were analysed by anti-Myc (upper) and anti-GST (lower left) immunoblot. **b**, The kinase activity of Myc-Raf (WT), myc-Raf(S259A) or Myc-Raf(S259A/S621A) was measured after anti-Myc immunoprecipitation, carried out directly (left panel) or after incubation (for 60 min) of extracts with vehicle, 0.3 mM phosphopeptide (PP) or unphosphorylated peptide (P) (right panel). **c**, GST-Raf, transfected together with Myc/14-3-3, Myc-cRaf-1 (WT), Myc-Raf (S259A) or Myc-Raf (S259A/S621A), was purified and analysed by anti-Myc (upper left) and anti-GST (lower left)

immunoblot. The initial extracts were also subjected to anti-Myc immunoblot (upper right). **d**, Baculoviral recombinant GST-Raf and His₆-Raf were expressed alone or together. Extracts of doubly infected cells were serially diluted ('NO' means no dilution) and incubated with vehicle, or with 1 mM peptides (P or PP). After 3 h, GST-Raf polypeptides were purified and immunoblotted with anti-Raf antibodies. **e**, Left panel, COS cells expressing an FKBP/haemagglutinin (HA)/Raf fusion protein were treated with carrier, EGF, the dimerizing agent FK1012A or both EGF and FK1012A. Extracts were incubated for 60 min with vehicle or 0.3 mM peptide (P or PP). Anti-HA immunoprecipitates were assayed for Raf kinase activity. Right panel, we studied the ability of phosphopeptide to displace FKBP/HA/Raf from co-expressed GST/14-3-3 *in vitro*.

Readdition of different types of 14-3-3, after phosphopeptide removal, restored substantial Raf activity (Fig. 2a), further indicating that phosphopeptide-induced inhibition of Raf is due to simple displacement of 14-3-3. Purified bacterial recombinant 14-3-3 added directly to phosphopeptide-inactivated GST-Raf immobilized on glutathione (GSH)-Sephacrose is sufficient to restore Raf kinase activity to ~40–50% of initial levels. The 14-3-3 protein can restore activity to Raf that has been inactivated by phosphopeptide *in vitro* after having previously been activated *in vivo*. However, 14-3-3 addition *in vitro* does not confer activity on GST-Raf (with a mutation of lysine 375 to methionine) isolated from EGF-treated cells (Fig. 2a), and nor does 14-3-3 alter the activity of previously inactive Raf, isolated from serum-deprived COS cells, from which 14-3-3 was displaced *in vitro* using synthetic phosphopeptide (Fig. 2b). The ability of prokaryotic recombinant 14-3-3 to reactivate Raf reflects the phosphoserine-specific binding of 14-3-3 to Raf, because treatment of Raf with protein phosphatase-1, after phosphopeptide-induced displacement of 14-3-3, abolishes completely the ability of 14-3-3 to reactivate Raf (Fig. 2c). Thus, Raf activity requires a continuous, phosphoserine-specific association with 14-3-3. This requirement is independent of the ability of 14-3-3 to protect Raf from dephosphorylation and to bring another polypeptide (other than Raf) into the complex.

We next determined whether the ability of 14-3-3 to maintain and/or restore Raf activity *in vitro* is attributable to an ability of 14-3-3 to dimerize Raf. The phosphoserine-dependent binding of 14-3-3 to Raf is mediated by Ser 259 and Ser 621 of Raf^{7,19}. Mutation of either Ser 259 or Ser 621 to Ala diminishes the association of 14-3-3 and Raf *in vivo*¹⁹ and a double mutant (with mutation of both Ser 259 and Ser 621 to Ala) exhibits no 14-3-3 binding, whether

examined in a yeast two-hybrid assay (Table 1) or by transfection of mutant Raf with 14-3-3 into mammalian cells (Fig. 3a). Mutation of Ser 259 to Ala results in a modest increase in basal Raf activity, and allows a robust activation in response to EGF that is consistently greater than that seen with wild-type Raf (see also refs 9, 19, 23). Nevertheless, the activity of Raf(S259A), like that of wild-type Raf, is also deactivated by phosphopeptide (Fig. 3b, right). The Raf(S621A) mutant and the double mutant are inactive (Fig. 3b). The level of expression of the double mutant is well maintained, as compared with wild-type Raf, on the basis of both its unimpaired ability to associate with V12Ras and MEK in the yeast two-hybrid assay (Table 1) and a comparison of expression of wild-type and mutant Raf polypeptide in COS cells.

Despite the nearly complete loss of 14-3-3 binding by the Raf double mutant, this mutant can still form homo-oligomers *in vivo* in COS cells (Fig. 3c). Moreover, Raf-Raf homo-oligomers formed during expression of Raf in Sf9 cells are unaffected by addition of phosphopeptide at concentrations sufficient almost to completely deactivate Raf (compare Figs 1d and 3d). Thus, Raf oligomerization *in vivo* is probably not mediated by 14-3-3. The activation *in vivo* of an FKBP-Raf fusion protein by FK1012-induced oligomerization²⁴, like the activation by EGF, depends on the binding of 14-3-3 to the FKBP-Raf polypeptide, as FK1012-activated FKBP-Raf is susceptible to inhibition by incubation *in vitro* with phosphopeptide before immunoprecipitation (Fig. 3e). Thus, forced oligomerization of Raf does not overcome the deactivation caused by phosphopeptide-induced displacement of 14-3-3.

Although the ability of 14-3-3 to cause Raf reactivation *in vitro* does not depend on its ability to promote Raf dimerization, 14-3-3 itself must be dimeric. In addition to the drastic N-terminal

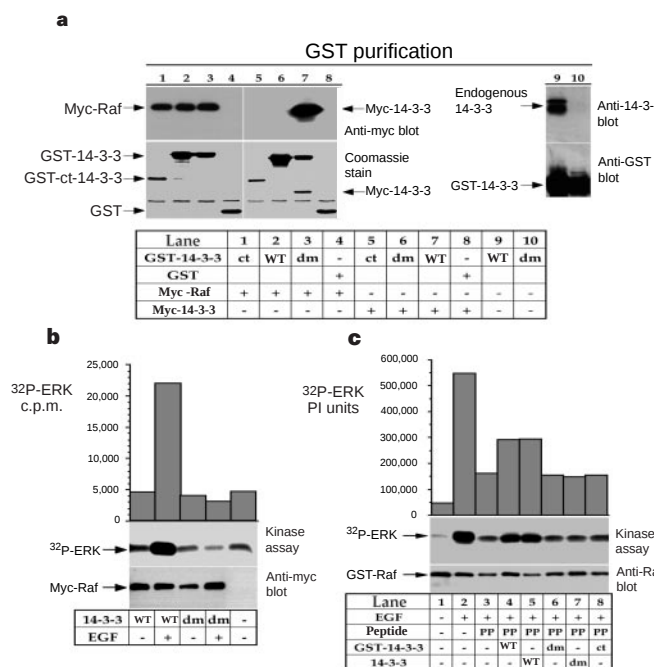
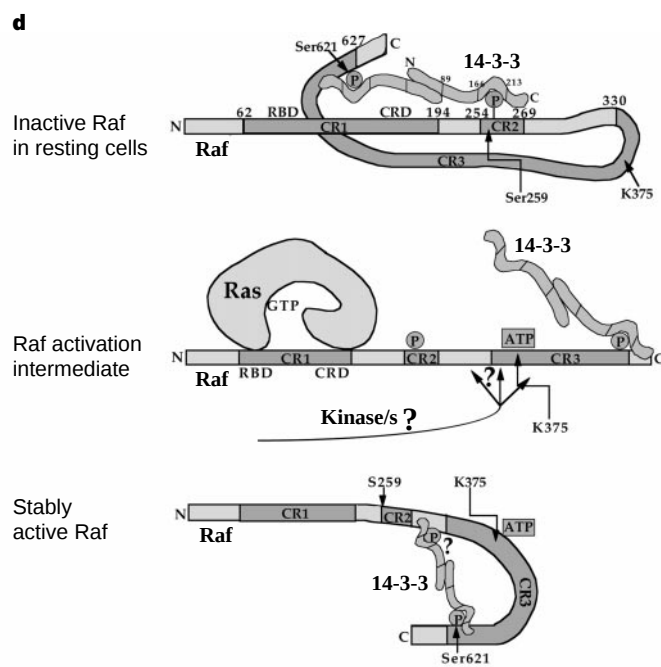


Figure 4 Dimerization-deficient mutants of 14-3-3 bind Raf but do not support Raf activity *in vivo* or *in vitro*. **a**, Wild-type GST-14-3-3 (WT), GST/14-3-3 dimer mutant (dm), GST/14-3-3(139–245) (a monomeric N-terminal-truncation mutant; ct) or GST was transfected alone or with Myc-Raf or Myc/14-3-3. The purified GST-fusion proteins were analysed for binding *in vivo* to Myc-Raf, Myc/14-3-3 and cellular 14-3-3. **b**, COS cells coexpressing Myc-Raf together with GST/14-3-3 (WT) or GST/14-3-3 dimer mutant (dm) were treated with vehicle or EGF. The purified, immobilized GST/14-3-3 proteins were released with GSH (20 mM, pH 8). GST/14-3-3-associated Myc-Raf was isolated by anti-Myc immunoprecipitation and subjected to the Raf kinase assay and immunoblotting with anti-Myc antibodies. **c**, COS cells expressing GST-Raf were treated with carrier or EGF. The extracts were incubated with vehicle or 0.3 mM phosphopeptide (PP) for 60 min. After

truncation used previously¹⁸ (Fig. 4a, GST-14-3-3(139–245); compare lanes 1, 2), we have also used site-specific mutation within the 14-3-3 dimerization interface²⁶ to engineer a monomeric 14-3-3 (Fig. 4a: compare lanes 6, 7, 9, 10). The dimerization-deficient 14-3-3 polypeptide continues to bind Myc-epitope-tagged Raf as well as does wild-type 14-3-3 (Fig. 4a: compare lanes 2, 3); however, the Raf polypeptide recovered with the dimer-deficient 14-3-3 is inactive (Fig. 4b). This indicates that monomeric 14-3-3 cannot initiate and/or maintain the activation of Raf *in vivo*. To determine whether monomeric 14-3-3 is deficient only in the initiation of Raf activation, we compared the ability of monomeric and dimeric 14-3-3 polypeptides to reactivate Raf *in vitro* after displacement of endogenous 14-3-3. About 40% recovery of Raf activity was obtained after addition of wild-type 14-3-3 polypeptide, whereas dimer-deficient 14-3-3 failed to restore Raf activity (Fig. 4c).

Our results show that Raf activation requires formation of a complex containing Raf and 14-3-3, and clarify the role of 14-3-3 in Raf activation. The requirement for dimeric 14-3-3 indicates that 14-3-3 reactivates Raf *in vitro* either by creating a Raf-Raf homodimer or by binding to a single Raf polypeptide at two sites. Evidence against 14-3-3-induced Raf-Raf dimerization in Raf activation is (1) the inability of drug-induced oligomerization of FKBP-Raf to sustain Raf catalytic activity if endogenous 14-3-3 is displaced *in vitro*; and (2) the inability of prokaryotic 14-3-3 to activate *in vitro* an initially inactive Raf polypeptide obtained from serum-deprived cells and stripped of 14-3-3 *in vitro*. Some modification of Raf occurs during activation *in vivo*, so that the binding of 14-3-3 *in vitro* to Raf previously activated *in vivo*



purification, immobilized GST-Raf was incubated with buffer, bacterial recombinant wild-type GST/14-3-3, recombinant wild-type bacterial 14-3-3 (thrombin-cleaved, lacking GST), GST/14-3-3 dimer mutant (thrombin-cleaved), or GST/14-3-3(139–245)(ct). After 90 min the GST-Raf polypeptides were washed and assayed for kinase activity. **d**, A model of the role of 14-3-3 in the Ras-dependent activation of cRaf-1. Top, Raf is maintained in an inactive state by the binding of a 14-3-3 dimer at two sites (Ser259-P and Ser621-P) on a single Raf polypeptide. Centre, half of the 14-3-3 dimer is displaced from Raf Ser259-P by Ras-GTP. Phosphorylation of further sites on Raf may occur (this phosphorylation can be stimulated by EGF, for example). Bottom, the 14-3-3 dimer has bound to a newly phosphorylated site on Raf, stabilizing an active conformation of Raf.

generates a conformation that is different from the one generated by binding of 14-3-3 to the Raf that was inactive *in vivo*.

We propose the following model for the role of 14-3-3 in the regulation of Raf activity (Fig. 4d). Raf is maintained in an inactive state by the binding of a 14-3-3 dimer to a single Raf polypeptide at two sites, phosphorylated Ser 259 and Ser 621. Ras-GTP displaces 14-3-3 from phosphorylated Ser 259 (ref. 27). As predicted by the results shown in Fig. 2c, we detect (as-yet unidentified) EGF-stimulated phosphorylation of Raf on sites that are protected from phosphatase treatment *in vitro* in the presence of 14-3-3 (data not shown); thus these may include a new 14-3-3-binding site. We propose that half of the 14-3-3 dimer displaced from phosphorylated Ser 259 by Ras-GTP rebinds to this putative new phosphoserine site, thereby stabilizing an active conformation that no longer requires Ras-GTP. This model incorporates most of the available genetic and biochemical evidence and provides a new paradigm for the function of 14-3-3, which both inhibits unstimulated Raf kinase and enables Raf to be stably activated by Ras-GTP and an unidentified (serine/threonine) kinase that acts on cRaf-1. □

Methods

Raf and 14-3-3 complementary DNA constructions. The Raf(S259A) mutant and double mutant were created using site-directed mutagenesis in the pBJ vector and transferred as an EcoRI fragment to the pMT2-Myc vector¹⁸. The 14-3-3 dimerization-deficient mutant was created using site-directed mutagenesis of the following 14-3-3 amino acids: E5K, L12AE to Q12QR, Y82Q, K85N and E87Q²⁶. Mutagenesis was carried out in the pGEX vector; the mutant 14-3-3 was subcloned into the vector pEBG¹⁸ as a BamHI/EcoRV

fragment for expression in mammalian cells. GST-Raf, GST/14-3-3 ζ (residues 1–245), GST/14-3-3(1–140) and GST/14-3-3(139–245) were expressed in mammalian cells using the vector pEBG. The construction of these cDNAs¹⁸ and the construction of FKBP/haemagglutinin/Raf cDNA²⁴ are described elsewhere.

Preparation of recombinant proteins. MEK1, ERK-2(KR), the wild-type 14-3-3 protein (residues 1–245), the 14-3-3 dimer mutant, or 14-3-3 polypeptides consisting of residues 1–140 or 139–245 were expressed in *Escherichia coli* as GST-fusion proteins using the pGEX-KG vector. The proteins were purified using GSH–Sepharose and eluted with GSH (50 mM, pH 8). Alternatively, thrombin was added to the GSH–Sepharose-immobilized GST-fusion proteins, and the cleaved polypeptides were collected.

Cell culture and transfections. COS cells were transfected by the DEAE-dextran method using 5 μ g DNA per 10-cm plate or 12.5 μ g DNA per 15-cm plate. In co-transfections we used 2.5 μ g or 7.5 μ g of each plasmid for transfection. forty-eight hours after transfection, cells were deprived of serum for 24 h, and then treated with carrier or with EGF (Sigma) (100 ng ml⁻¹) for 20 min before collection.

Cell extraction and Raf kinase assay. Cells were extracted into ice-cold buffer (50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 1% Triton-X100, 1 mM dithiothreitol (DTT), 1 mM EDTA, 1 mM EGTA, 2 mM Na₃VO₄, 50 mM β -glycerophosphate and proteinase inhibitors). GST-fusion proteins were purified by adsorption to GSH–Sepharose (Pharmacia) at 4 °C. Immunoprecipitations were carried out at 4 °C with the appropriate antibodies prebound to protein A/G agarose (Santa Cruz). After protein adsorption, the GSH–Sepharose or protein A/G agarose beads were washed twice with extraction buffer, twice with extraction buffer containing 0.5 M LiCl and twice with kinase assay buffer (40 mM Tris-Cl, pH 7.5, 0.1 mM EDTA, 5 mM MgCl₂ and 2 mM DTT). The Raf kinase assay contained (~100 μ l final volume) the washed beads in kinase assay buffer, 100 μ M ATP, 10 μ Ci [γ -³²P]ATP and 0.3 μ g recombinant GST–MEK-1. After incubation at 30 °C for 20 min, 2 μ g bacterial recombinant inactive ERK-2(KR) was added, and incubation was continued for another 30 min. Assay blanks lacked Raf only, and are shown. After SDS–PAGE and electrophoretic transfer to PVDF membranes, ³²P incorporation into ERK-2(KR) was quantified either by Cerenkov counting of the excised ³²P-ERK band or by Phosphor-Imager (PI) analysis; data are presented as c.p.m. or PI units.

Dissociation of the 14-3-3/Raf complex in vitro. The phosphopeptide (called PP, sequence LPKINRSA(Sp)EPSLHR, corresponding to cRaf-1 amino acids 613–627; Sp is serine phosphate) or the same peptide in an unphosphorylated form (called P) was added to cell extracts at concentrations between 0.3 and 1.0 mM. Incubations were at 4 °C for 30–180 min. We used a mixture of phosphatase inhibitors (50 mM β -glycerophosphate, 10 mM sodium pyrophosphate, 10 mM NaF, 25 mM *p*-nitrophenylpyrophosphate, 0.3 mM calyculin-A and 2 mM Na₂VO₃) in some experiments. Incubations were terminated by adsorption to GSH–Sepharose or by immunoprecipitation, and this termination was followed by washing.

Raf reactivation in vitro. GST–Raf immobilized on GSH–Sepharose beads was incubated with extraction buffer containing 15 μ g GST/14-3-3 or an equal amount of the indicated protein in 1 ml, for 90 min at 4 °C. Incubations were terminated by washing as described for GSH–Sepharose purification. In some experiments, a cell extract was used in place of recombinant GST-fusion protein.

Two-hybrid assay. The vectors and yeast strains have been described²⁸. Protein–protein interactions are indicated by the expression of LacZ, detected by the development of a blue colour in a filter lift assay using X-gal²⁸.

Received 29 December 1997; accepted 28 April 1998.

- Avruch, J., Zhang, X. F., Kyriakis, J. M. Ras meets Raf: completing the framework of a signal transduction pathway. *Trends Biochem. Sci.* **19**, 279–283 (1994).
- Stancato, L. F. et al. Raf exists in a native heterocomplex with hsp90 and hsp50 that can be reconstituted in a cell-free system. *J. Biol. Chem.* **268**, 21711–21716 (1993).
- Wartmann, M. & Davis, R. J. The native structure of the activated Raf protein kinase is a membrane-bound multi-subunit complex. *J. Biol. Chem.* **269**, 6695–6701 (1994).
- Aitken, A. 14-3-3 and its possible role in co-ordinating multiple signalling pathways. *Trends Cell Biol.* **6**, 341–347 (1996).
- Marshall, C. J. Ras Effectors. *Curr. Opin. Cell Biol.* **89**, 197–204 (1996).
- Zhang, X.-F. et al. Normal and oncogenic p21ras proteins bind to the amino-terminal regulatory domain of c-Raf-1. *Nature* **364**, 308–313 (1993).
- Stokoe, D. & McCormick, F. Activation of cRaf-1 by Ras and Src through different mechanisms: activation in vivo and in vitro. *EMBO J.* **16**, 2384–2396 (1997).
- Kyriakis, J. M., Force, T. L., Rapp, U. R., Bonventre, J. V. & Avruch, J. Mitogen regulation of c-Raf-1 protein kinase activity toward mitogen-activated protein kinase-kinase. *J. Biol. Chem.* **268**, 16009–16019 (1993).

- Morrison, D. K., Heidecker, G., Rapp, U. R. & Copeland, T. D. Identification of the major phosphorylation sites of the Raf-1 kinase. *J. Biol. Chem.* **268**, 17309–17316 (1993).
- Dent, P., Jelinek, T., Morrison, D. K., Weber, M. J. & Sturgill, T. W. Reversal of Raf-1 activation by purified and membrane-associated protein phosphatases. *Science* **268**, 1902–1906 (1995).
- Morrison, D. & Cutler, R. E. Jr The complexity of Raf-1 regulation. *Curr. Opin. Cell Biol.* **9**, 174–179 (1997).
- Chang, H. C. & Rubin, G. M. 14-3-3 ϵ positively regulates Ras-mediated signalling in *Drosophila*. *Genes Dev.* **11**, 1132–1139 (1997).
- Knockel, L., Vorbruggen, G., Jackle, H., Mlodzik, M. & Bohmann, D. Requirement for *Drosophila* 14-3-3 ζ in Raf-dependent photoreceptor development. *Genes Dev.* **11**, 1140–1147 (1997).
- Freed, E., Symons, M., Macdonald, S. G., McCormick, F. & Ruggieri, R. Binding of 14-3-3 proteins to the protein kinase Raf and effects on its activation. *Science* **265**, 1713–1716 (1994).
- Irie, K. et al. Stimulatory effects of yeast and mammalian 14-3-3 proteins on the Raf protein kinase. *Science* **265**, 1716–1719 (1994).
- Fantl, W. J. et al. Activation of Raf by 14-3-3 proteins. *Nature* **371**, 612–614 (1994).
- Muslin, A. J., Tanner, J. W., Allen, P. M. & Shaw, A. S. Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine. *Cell* **84**, 889–898 (1996).
- Luo, Z.-J., Zhang, X.-F., Rapp, U. & Avruch, J. Identification of the 14-3-3 ζ domains important for self-association and Raf binding. *J. Biol. Chem.* **270**, 23681–23687 (1995).
- Michaud, N. R., Fabian, J. R., Mathes, K. D. & Morrison, D. K. 14-3-3 is not essential for Raf-1 function: identification of cRaf-1 proteins that are biologically activated in a 14-3-3 and Ras-independent manner. *Mol. Cell Biol.* **15**, 3390–3397 (1995).
- Fu, H. et al. Interaction of the protein kinase Raf-1 with 14-3-3 proteins. *Science* **266**, 126–129 (1994).
- Suen, K.-L., Bustelo, X. R. & Barbacid, M. Lack of evidence for the activation of the Ras/Raf mitogenic pathway by 14-3-3 proteins in mammalian cells. *Oncogene* **11**, 825–831 (1995).
- Clark, G. J. et al. 14-3-3 ζ negatively regulates Raf-1 activity by interactions with the Raf-1 cysteine-rich domain. *J. Biol. Chem.* **272**, 20990–20993 (1997).
- Rommel, C., Radziwill, G., Moelling, K. & Hafen, E. Negative regulation of Raf activity by binding 14-3-3 to the aminoterminal of Raf in vivo. *Mech. Dev.* **4**, 95–104 (1997).
- Luo, Z.-J., Tzivion, G., Belshaw, P. J., Marshall, M. & Avruch, J. Oligomerization activates c-Raf-1 through a Ras-dependent mechanism. *Nature* **383**, 181–185 (1996).
- Farrar, M. A., Alberola-Ila & Perlmutter, R. M. Activation of the Raf-1 kinase cascade by coumermycin-induced dimerization. *Nature* **383**, 178–181 (1996).
- Liu, D. et al. Crystal structure of the zeta isoform of the 14-3-3 protein. *Nature* **376**, 191–194 (1995).
- Wan, C. et al. Activated Ras displaces 14-3-3 protein from the aminoterminal of cRaf-1. *Oncogene* **12**, 409–419 (1996).
- Farflee, T. et al. The Retinoblastoma protein associates with protein phosphatase type 1 catalytic subunit. *Genes Dev.* **7**, 555–569 (1993).

Acknowledgements. The authors were supported by a grant from Eli Lilly, Inc. We thank A. Khatri for peptide synthesis and purification, and J. Prendable for preparation of the manuscript.

Correspondence and requests for materials should be addressed to J. A. (e-mail: avruch@helix.mgh.harvard.edu).

The oncoprotein Evi-1 represses TGF- β signalling by inhibiting Smad3

Mineo Kurokawa*, Kinuko Mitani*, Kenji Irie†, Tomohiro Matsuyama*, Tokiharu Takahashi*, Shigeru Chiba*, Yoshio Yazaki*, Kunihiro Matsumoto† & Hisamaru Hirai*

* The Third Department of Internal Medicine, Faculty of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

† Department of Molecular Biology, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan

Evi-1 encodes a zinc-finger protein that may be involved in leukaemic transformation of haematopoietic cells^{1–5}. Evi-1 has two zinc-finger domains, one with seven repeats of a zinc-finger motif and one with three repeats⁶, and it has characteristics of a transcriptional regulator^{7,8}. Although Evi-1 is thought to be able to promote growth and to block differentiation in some cell types^{9–11}, its biological functions are poorly understood. Here we study the mechanisms that underlie oncogenesis induced by Evi-1 by investigating whether Evi-1 perturbs signalling through transforming growth factor- β (TGF- β), one of the most studied growth-regulatory factors, which inhibits proliferation of a wide range of cell types¹². We show that Evi-1 represses TGF- β signalling and antagonizes the growth-inhibitory effects of TGF- β . Two separate regions of Evi-1 are responsible for this repression; one of these regions is the first zinc-finger domain. Through this domain, Evi-1 interacts with Smad3, an intracellular mediator of TGF- β signalling¹³, thereby suppressing the transcriptional activity of Smad3. These results define a new function of Evi-1 as a repressor of signalling through TGF- β .

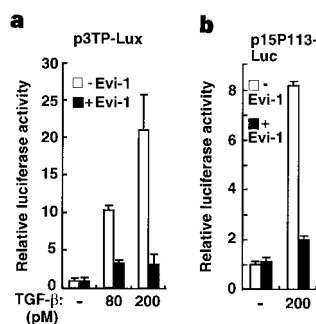


Figure 1 TGF β -mediated transcriptional responses are suppressed by Evi-1. **a**, p3TP-Lux, or, **b**, p15P113-Luc was transfected into HepG2 cells together with either pME18S or pME-Evi-1, and relative luciferase activity was measured. Values of relative luciferase activity and error bars represent the means and the standard deviations, respectively, of four separate experiments.

We examined the effects of Evi-1 on TGF β -mediated transcriptional responses using co-transfection assays. We first used p3TP-Lux, a TGF β -responsive reporter plasmid consisting of the plasminogen-activator inhibitor-1 (PAI-1) promoter¹⁴. When p3TP-Lux alone was transfected, a significant increase in luciferase activity was observed in the presence of TGF- β (Fig. 1a)¹⁵. These transactivations were repressed almost to control levels when Evi-1 was transfected with p3TP-Lux. Similar repression occurred when we used another TGF β -responsive reporter that contained the p15 promoter¹⁶ (Fig. 1b). We performed these experiments using

Mv1Lu cells, and similar results were obtained (data not shown). There was no effect of Evi-1 on the basal activity of these promoters, indicating that these Evi-1 effects are specific to TGF β -induced transcriptional activation. None of the consensus DNA sequences for Evi-1 binding is found in the promoters studied here, indicating that Evi-1 may repress TGF- β signalling by interrupting intracellular signalling pathways, rather than binding directly to the target promoters.

To determine whether Evi-1 could affect the antiproliferative effects of TGF- β *in vivo*, we established several Mv1Lu cell lines that stably express Evi-1 (Fig. 2a). When cultured in the absence of TGF- β , all of these clones showed similar proliferative abilities. In the presence of TGF- β , the growth of the control cell lines, M-2 and M-4, was effectively inhibited (Fig. 2b). In contrast, overexpression of Evi-1 prevented the cells from undergoing growth arrest even after 72 h exposure to TGF- β . [³H]thymidine-incorporation assays showed that the growth of the M-2 and M-4 lines was inhibited by 0.3 pM TGF- β , whereas overexpression of Evi-1 in E-5, E-31 and E-82 cell lines markedly diminished the responsiveness to TGF- β in an expression-level-dependent manner (Fig. 2c). We assayed the state of phosphorylation of the retinoblastoma protein (Rb) as an *in vivo* measurement of Evi-1 effects. We found most of the Rb in the Evi1-expressing cells remained in the hyperphosphorylated state even in the presence of TGF- β , whereas the control clones treated with TGF- β accumulated hypophosphorylated Rb and showed a decrease in the overall level of detectable Rb (Fig. 2d)¹⁷.

To identify the target through which Evi-1 represses TGF- β signalling, we tested whether Evi-1 could interact with the SMAD proteins, signalling components required for TGF- β signalling^{18–20}. We studied Smad 1–4; Evi-1 was specifically detected in the immune complex of Smad3 (Fig. 3a). These results suggest a

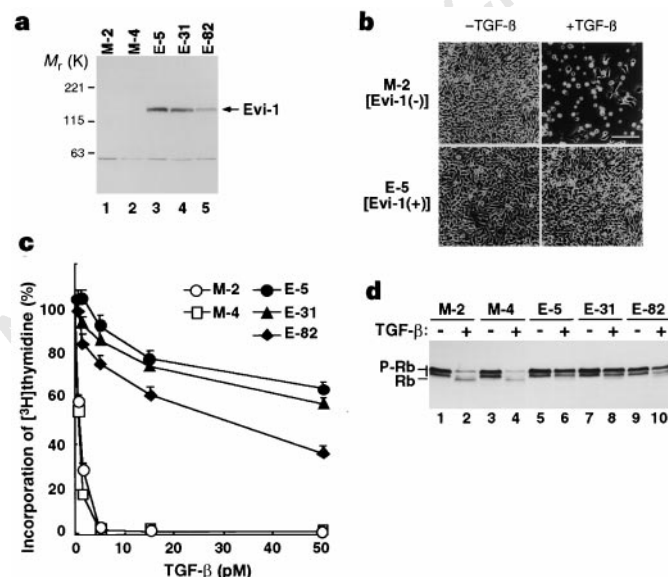


Figure 2 Constitutive expression of Evi-1 in Mv1Lu cells overcomes TGF β -mediated growth inhibition. **a**, Expression of Evi-1 in Mv1Lu clones. Clones M-2 and M-4 were transfected with pME18Sneo. Clones E-5, E-31 and E-82 were established from cells transfected with Evi-1. **b**, The Mv1Lu clones were cultured in the absence or presence of 20 pM TGF- β . Figures show morphologies of the representative clones after a 72-h incubation. Scale bar represents 250 μ m. **c**, [³H]thymidine incorporation into Mv1Lu clones was assayed in the presence of the indicated concentrations of TGF- β . Results are expressed as percentages of the values obtained from control cultures that did not receive TGF- β . **d**, Exponentially growing Mv1Lu clones were cultured in the absence or presence of 50 pM TGF- β for 12 h. Western blot analysis of Rb was performed using whole-cell extracts. The positions of hyperphosphorylated (P-Rb) and hypophosphorylated (Rb) forms of Rb are shown.

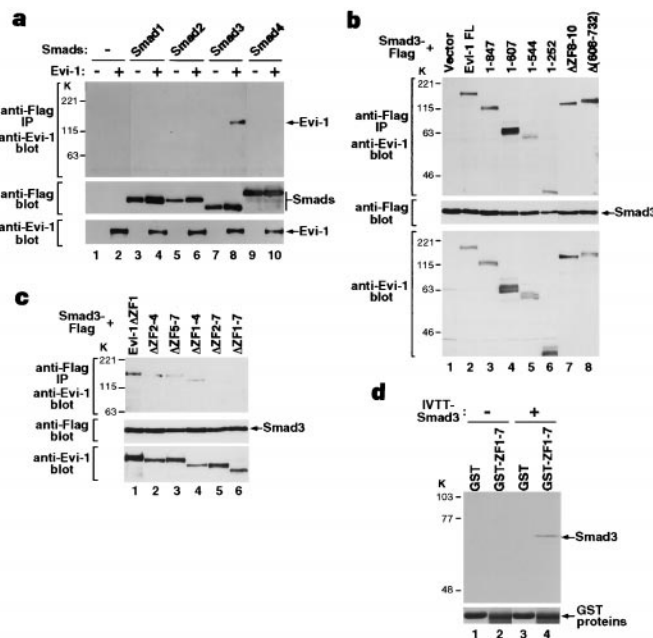


Figure 3 Evi-1 associates with Smad3 through the first zinc-finger domain of Evi-1. **a**, Evi-1 was transfected into COS7 cells with the indicated Flag-tagged Smad constructs. Cell extracts were subjected to immunoprecipitation (IP) with anti-Flag. This was followed by immunoblotting with anti-Evi-1 antibodies. Expression of SMADs and Evi-1 was monitored. **b**, **c**, The indicated forms of Evi-1 were transfected into COS7 cells together with Smad3-Flag. Cell extracts were immunoprecipitated with anti-Flag antibodies and then immunoblotted with anti-Evi-1 antibodies. Expression of Smad3-Flag and the Evi-1 mutants is shown. FL, full length. **d**, Binding of [³⁵S]methionine-labelled Smad3 that was synthesized *in vitro* (IVTT-Smad3) to GST/Evi-1ZFI-7. GST fusion proteins were also visualized.

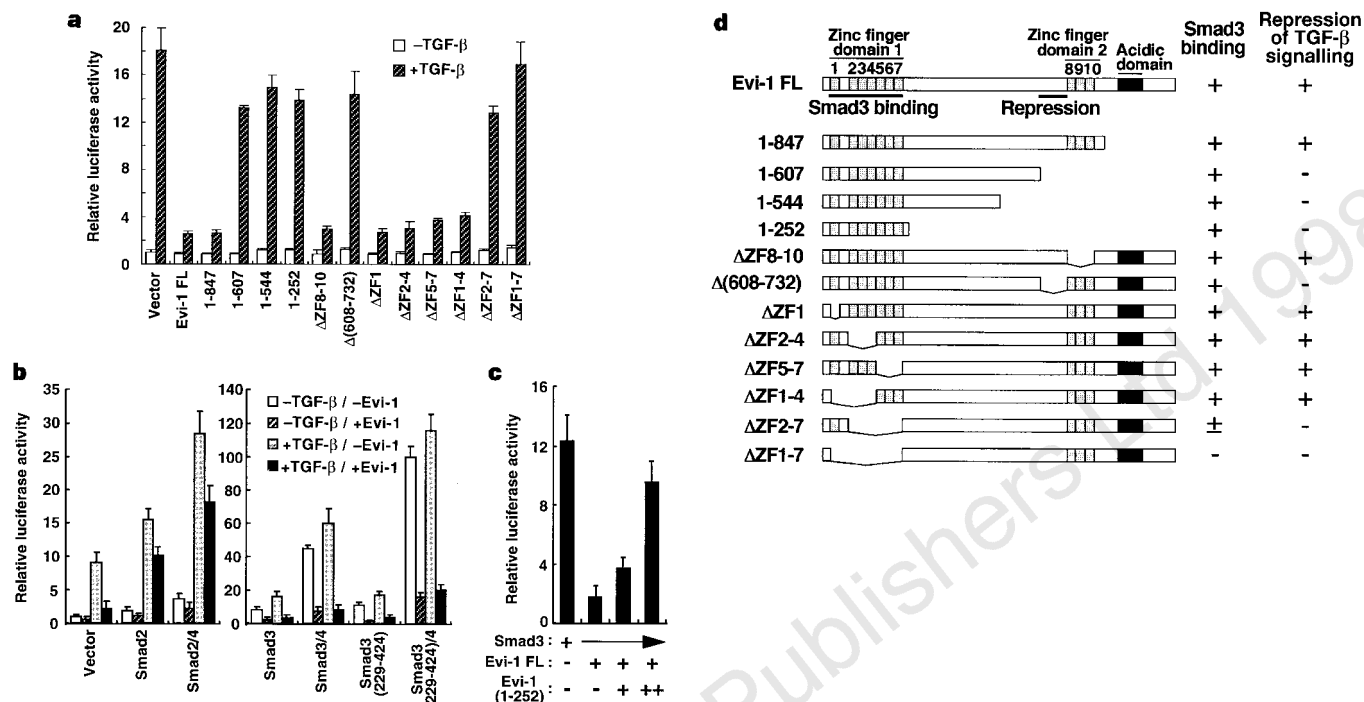


Figure 4 Contribution of Evi-1 domains to the repression of TGF- β and SMAD signalling. **a**, Structural requirements for repression of TGF- β signalling by Evi-1. p3TP-Lux was transfected into HepG2 cells with each mutant of Evi-1 in the absence or the presence of 200 pM TGF- β . Relative luciferase activity was measured. FL, full length. **b**, Evi-1 inhibits Smad3-induced responses to TGF- β . Either pME18S or pME-Evi-1 in combination with p3TP-Lux was transfected into HepG2 cells together with the indicated constructs in the absence or the

presence of 200 pM TGF- β . **c**, Evi-1(1-252) rescues the Smad3 function that is suppressed by full-length Evi-1. Smad3 and p3TP-Lux were transfected together into HepG2 cells with 0.5 μ g of pME-Evi-1 FL and increasing amounts (+, 0.6 μ g; ++, 3 μ g) of pME-Evi-1(1-252). **d**, Structures of full-length Evi-1 and of its deletion mutants. Regions necessary for association with Smad3 or repression of TGF- β signalling are shown. Zinc-finger motifs are numbered 1-10.

model in which Evi-1 antagonizes TGF- β signalling by quenching Smad3. Using metabolic labelling, we estimated that ~5% of immunoprecipitable Smad3 forms a complex with Evi-1 when both are expressed in COS7 cells (data not shown). To locate the Smad3-binding domain of Evi-1, we evaluated the abilities of a series of specific deletion mutants of Evi-1 to bind Smad3. The structures of these deletion mutants are shown in Fig. 4d. As shown in Fig. 3b, Evi-1(1-252), which consists mainly of the first zinc-finger domain, retained the ability to bind Smad3. Detailed internal deletions in the first zinc-finger domain showed that the deletion including the entire first zinc-finger domain (ZF1-7) completely abolishes Evi-1-Smad3 binding (Fig. 3c). Thus, the first zinc-finger domain of Evi-1 is essential for the interaction between Evi-1 and Smad3. This interaction is thought to be direct, as shown by a pull-down assay (Fig. 3d). All SMAD proteins share structural conservation of an amino-terminal domain and a carboxy-terminal domain, designated MH1 and MH2, respectively¹⁸. Using co-immunoprecipitation assays, we found that Smad3 interacts with Evi-1 through the MH2 domain (data not shown).

Smad3 is directly phosphorylated by the activated TGF- β type I receptor (T β RI), associates with Smad4, and is subsequently translocated into the nucleus^{13,21,22}. To determine which of these processes is affected by Evi-1, we first examined ligand-dependent phosphorylation of Smad3. As shown in Fig. 5A, Smad3 is phosphorylated by the constitutively activated T β RI (T β RI TD)²³ in the presence of Evi-1 as strongly as it is in the absence of Evi-1, indicating that Evi-1 does not inhibit receptor-dependent phosphorylation of Smad3. The ligand-induced formation of heteromers containing Smad3 and Smad4 was not disrupted by the presence of Evi-1 (data not shown). As shown in Fig. 5B, no significant change in the TGF β -induced nuclear translocation of Smad3 was observed in the presence of Evi-1, and Evi-1 was consistently nuclear regardless of

TGF- β stimulation. These results indicate that Evi-1 probably does not block the nuclear accumulation of Smad3 and that Evi-1 may interact in the nucleus with the receptor-activated Smad3-Smad4 (Smad3/4) complex.

Consistent with this model, TGF- β stimulation significantly increases the amount of the Evi-1-Smad3 complex (Fig. 5C), whereas Evi-1 cannot be detected in the immune complex of Smad2 even after stimulation with TGF- β . As shown in Fig. 3a, Evi-1 does not form a stable complex with Smad4 when these proteins are expressed together in COS7 cells. However, Evi-1 did immunoprecipitate with Smad4 when they were co-transfected with Smad3 and stimulated with TGF- β , indicating that a ternary complex containing Evi-1, Smad3 and Smad4 may form. We detected this ternary complex when cells were co-transfected with all three constructs in the presence of TGF- β (Fig. 5D). The association between Evi-1 and Smad3 was not affected by formation of the stable Smad3/4 complex (data not shown). We also observed that Evi-1 can interact equally with phosphorylated and dephosphorylated Smad3 (data not shown), indicating that Smad3 phosphorylation by itself does not alter the affinity of Evi-1 for Smad3. It has been reported that the Smad3/4 complex specifically recognizes a binding site within the PAI-1 promoter in p3TP-Lux²⁴. Therefore, we investigated whether Evi-1 affects the DNA-binding ability of the Smad3/4 complex by using an electrophoretic mobility shift assay with a probe derived from p3TP-Lux. As shown in Fig. 5E, levels of the Smad3/4-DNA complex were significantly decreased when Evi-1 was expressed with Smad3 and Smad4. Taken together, these data indicate that Evi-1 abrogates the interaction of the Smad3 complex with its binding sequence in DNA in the nucleus.

To evaluate the relevance of the Evi-1-Smad3 interaction to the repression of TGF- β signalling, we mapped the region of Evi-1 that is necessary for the repression. We used the C-terminally truncated

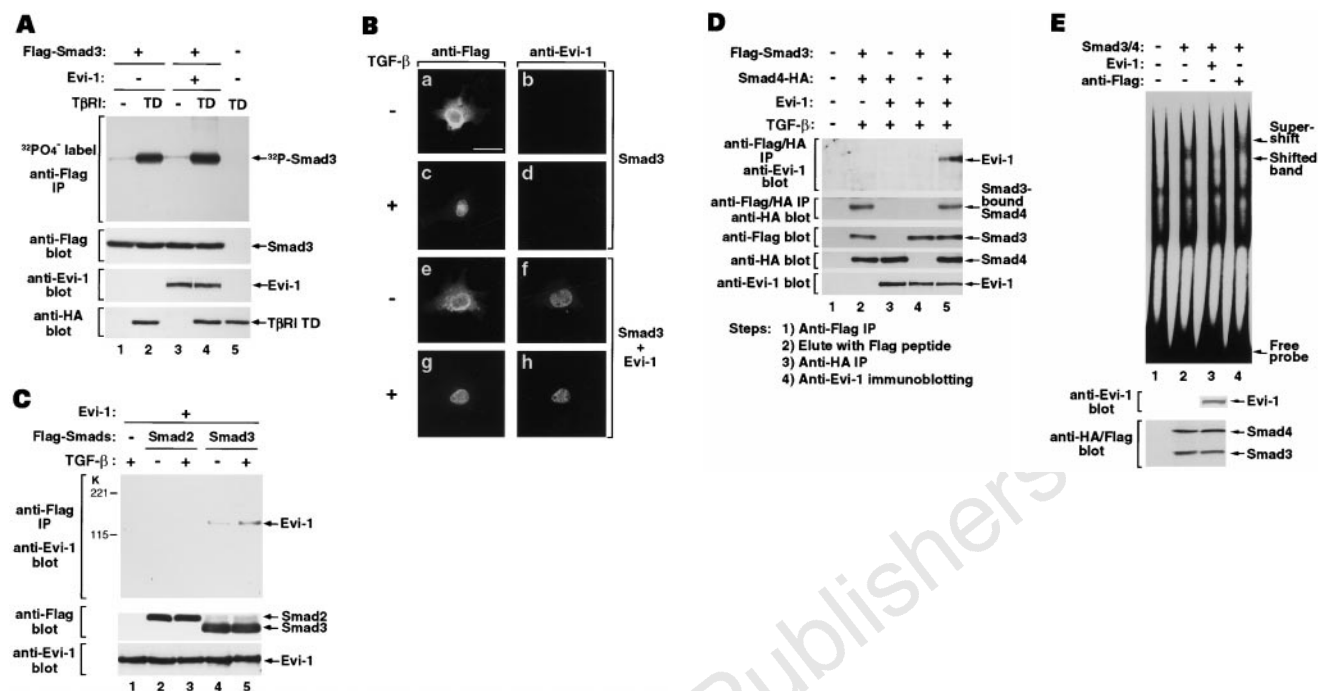


Figure 5 Evi-1 acts as a nuclear inhibitor of Smad3. **A**, Smad3 is phosphorylated by T β RI TD in the presence and absence of Evi-1. COS7 cells were transfected with the indicated combinations of Flag-Smad3, Evi-1 and T β RI TD. Cells were labelled with [32 P]phosphate and Flag-Smad3 was purified by immunoprecipitation (IP) with anti-Flag antibodies. The expression of each construct is presented. **B**, Evi-1 does not affect the nuclear localization of Smad3. COS7 cells were transfected with Flag-Smad3 alone (**a–d**) or in conjunction with Evi-1 (**e–h**) and were incubated either in the absence (**a, b, e, f**) or the presence (**c, d, g, h**) of TGF- β . Flag-Smad3 was visualized with anti-Flag antibodies, and Evi-1 was detected with anti-Evi-1 antibodies. Scale bar, 50 μ m. **C**, Stimulation by TGF- β increases the amount of the Evi-1-Smad3 complex. Cells were transfected with the indicated constructs and stimulated with TGF- β . Interactions between Evi-1 and Smads were analysed. Expression of Flag-SMADs and Evi-1 was monitored. **D**,

Detection of a ternary complex containing Evi-1, Smad3 and Smad4 in the presence of TGF- β . COS7 cells were co-transfected with the indicated constructs and stimulated with TGF- β . Cell extracts were immunoprecipitated with the agarose-immobilized anti-Flag antibody. Beads were eluted with Flag peptide and the eluate was precipitated with anti-HA antibodies. This step was followed by anti-Evi-1 immunoblotting. Expression of each construct was monitored. **E**, Levels of the Smad3/4-DNA complex are decreased in the presence of Evi-1. 293 cells were transfected with Flag-Smad3 and Smad4-HA in the absence or presence of Evi-1. Nuclear extracts were subjected to EMSA; the 60-base-pair oligonucleotide derived from the PAI-1 promoter sequence of p3TP-Lux was used as a probe. Shifted bands and a supershifted band labelled with anti-Flag antibodies are indicated. Evi-1 and Smad3/4 in the nuclear extracts were monitored.

versions of Evi-1 shown in Fig. 4d and demonstrated that the repressor activity resides in two separate regions, namely the region between amino acids 608 and 732 and the first zinc-finger domain (Fig. 4a). Amino acids 608–732 are unnecessary for Smad3 binding (Fig. 3b, lane 8). Therefore, the region containing amino acids 608–732 and the Smad3-binding domain are both required for the efficient repression of TGF- β signalling. These results also indicate that the known ability of Evi-1 to transform fibroblasts⁹ and to block myeloid-cell differentiation¹⁰ should be independent of repression of TGF- β signalling because these activities depend on the second zinc-finger domain^{9,25}.

To investigate the effects of Evi-1 on Smad signalling, we assayed transcriptional responses in the presence of various Smad proteins. As shown in Fig. 4b, Evi-1 fully inhibits Smad3- or Smad3/4-induced transcriptional activation. Smad3(229–424), which corresponds to the MH2 domain, activates the PAI-1 promoter, and this activation was also suppressed by Evi-1. Smad2 or a combination of Smad2 and Smad4 (Smad2/4) can also activate the PAI-1 promoter^{26,27}. In contrast to the case with Smad3, however, Evi-1 repressed the Smad2- or Smad2/4-induced responses only to 70–80% of the full activation (Fig. 4b). These lesser repressive effects of Evi-1 on Smad2 may reflect the inability of Evi-1 to interact with Smad2, and indicate that repression of TGF- β signalling by Evi-1 is dependent on Smad3 binding. Furthermore, Evi-1(1–252), which binds to Smad3 but does not block TGF- β signalling, can rescue the Smad3 function suppressed by the full-length Evi-1 in a dose-dependent manner (Fig. 4c). We conclude that Evi-1 is an inacti-

vator of Smad3 and that repression of TGF- β signalling by Evi-1 depends on the interaction of Evi-1 with Smad3.

We have shown that Evi-1 negatively regulates TGF- β signalling and that Evi-1 associates with Smad3 and that it interrupts DNA binding by the Smad3 complex. Identification of the interaction between two distinct transcriptional regulators has broad implications for the repertoire and complexity of the regulatory mechanisms underlying TGF- β and SMAD signalling, and provides insights into both physiological and oncogenic activities of the Evi-1 oncoprotein. □

Methods

Construction of expression vectors. The human Evi-1 complementary DNA was inserted into the *Eco*RI site of plasmid pME18S (pME-Evi-1) or pME18Sneo^{7,28}. C-terminally truncated versions of Evi-1 were made by digesting the full-length Evi-1 with appropriate enzymes. Construction of Evi-1 Δ ZF1, Δ ZF2–4, Δ ZF5–7 and Δ ZF8–10 has been described⁷. The other mutants of the first zinc-finger domain were generated from Evi-1 Δ ZF1, Evi-1 Δ ZF2–7 and Δ ZF5–7. Evi-1 Δ (608–732) was constructed using a polymerase chain reaction (PCR) method. Glutathione S-transferase (GST)/Evi-1ZF1–7 was made by inserting the corresponding fragments into pGEX-2TK (Pharmacia). All the SMAD fragments were subcloned into pCMV5. For Smad3(229–424)-Flag and Smad3-Flag, the corresponding fragments were generated by a PCR method.

Growth-inhibition assays. The Mv1Lu cell clones were incubated for 24 h in the presence of TGF- β . During the last 2 h, cells were labelled with 1 μ Ci ml⁻¹ [3 H]thymidine (Amersham). Cells were washed, fixed, and solubilized with

0.5 M NaOH. The cell extracts were neutralized with HCl and ^3H radioactivity was measured in a liquid scintillation β -counter (Aloka).

Western blot analysis, immunoprecipitation, metabolic labelling, and pull-down assays. Cells indicated were lysed in TNE buffer²⁶. Whole-cell extracts or immunoprecipitates produced with anti-Flag, anti-haemagglutinin (HA) or anti-Evi-1 were visualized by immunoblotting with anti-Evi-1 (ref. 7), anti-Rb (Pharmingen), anti-Flag (M2, Eastman Kodak) or anti-HA (12CA5, Boehringer) antibodies. For ligand-induced association assays, COS7 cells transiently transfected with the indicated constructs were washed and incubated in the presence of 200 pM TGF- β for 2 h. [^{32}P]phosphate labelling was performed as described²⁶. For pull-down assays, 5 μg of GST fusion proteins were collected on glutathione-Sepharose beads (Pharmacia) and incubated with [^{35}S]methionine-labelled Smad3 that was synthesized using TNT-coupled reticulocyte lysate systems (Promega).

Electrophoretic mobility shift assays. EMSAs were performed as described²⁹, except that the oligonucleotide ATGAGTCAGACACCTCTGGCTGTCTGGA-AGGGATGAGTCAGACACCTCTG GCTTTCTGGA was used as a probe. For antibody supershift assays, 3 μg anti-Flag (M2) antibody was added to the reaction mixture.

Transcription-response assays. Transcriptional-response assays were performed as described⁷, except that we used LipofectAmine (Gibco BRL) instead of a calcium phosphate method. As an internal control of transfection efficiency, a plasmid expressing β -galactosidase was co-transfected. Cells were treated for 24 h with TGF- β ; relative luciferase activity was then measured in cell extracts using the luciferase assay system (Promega) and a luminometer (Lumat, Berthold) and was normalized to the β -galactosidase activity.

Immunofluorescence. Cells were treated with anti-Flag and anti-Evi-1 antibodies and then incubated with fluorescein isothiocyanate (FITC)-conjugated donkey anti-mouse IgG and Texas-red-conjugated donkey anti-rabbit IgG (Jackson Immunologicals), respectively, as secondary antibodies. Images were obtained using a confocal microscope (MRC-1024, Bio-Rad). The cells were counterstained with Hoechst 33258 to identify nuclei and evaluate transfection efficiency.

Received 25 March; accepted 27 April 1998.

- Lopingo, M. C. & Perkins, A. S. Molecular analysis of Evi1, a zinc finger oncogene involved in myeloid leukemia. *Cur. Top. Microbiol. Immunol.* **211**, 211–222 (1996).
- Morishita, K. *et al.* Activation of EVI1 gene expression in human acute myelogenous leukemias by translocations spanning 300–400 kilobases on chromosome band 3q26. *Proc. Natl Acad. Sci. USA* **89**, 3937–3941 (1992).
- Mitani, K. *et al.* Generation of the AML1-EVI-1 fusion gene in the t(3;21)(q26;q22) causes blastic crisis in chronic myelocytic leukemia. *EMBO J.* **13**, 504–510 (1994).
- Ogawa, S. *et al.* Structurally altered Evi-1 protein generated in the 3q21q26 syndrome. *Oncogene* **13**, 183–191 (1996).
- Ogawa, S. *et al.* Increased Evi-1 expression is frequently observed in blastic crisis of chronic myelocytic leukemia. *Leukemia* **10**, 788–794 (1996).
- Morishita, K. *et al.* Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell* **54**, 831–840 (1988).
- Tanaka, T. *et al.* Evi-1 raises AP-1 activity and stimulates c-fos promoter transactivation with dependence on the second zinc finger domain. *J. Biol. Chem.* **269**, 24020–24026 (1994).
- Bartholomew, C., Kilbey, A., Clark, A. M. & Walker, M. The Evi-1 proto-oncogene encodes a transcriptional repressor activity associated with transformation. *Oncogene* **14**, 569–577 (1997).
- Kurokawa, M. *et al.* The AML1/Evi-1 fusion protein in the t(3;21) translocation exhibits transforming activity on Rat1 fibroblasts with dependence on the Evi-1 sequence. *Oncogene* **11**, 833–840 (1995).
- Morishita, K., Parganas, E., Matsui, T. & Ihle, J. N. Expression of the Evi-1 zinc finger gene in 32Dc13 myeloid cells blocks granulocytic differentiation in response to granulocyte colony-stimulating factor. *Mol. Cell. Biol.* **12**, 183–189 (1992).
- Kreider, B. L., Orkin, S. H. & Ihle, J. N. Loss of erythropoietin responsiveness in erythroid progenitors due to expression of the Evi-1 myeloid-transforming gene. *Proc. Natl Acad. Sci. USA* **90**, 6454–6458 (1993).
- Massagué, J. The transforming growth factor- β family. *Annu. Rev. Cell Biol.* **6**, 597–641 (1990).
- Zhang, Y., Feng, X., We, R. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF- β response. *Nature* **383**, 168–172 (1996).
- Wrana, J. L., Attisano, L., Wieser, R., Ventura, E. & Massagué, J. Mechanism of activation of the TGF- β receptor. *Nature* **370**, 341–347 (1994).
- Macias-Silva, M. *et al.* MADR2 is a substrate of the TGF β receptor and its phosphorylation is required for nuclear accumulation and signaling. *Cell* **87**, 1215–1224 (1996).
- Hannon, G. J. & Beach, D. p15^{INK4B} is a potential effector of TGF- β -induced cell cycle arrest. *Nature* **371**, 257–261 (1994).
- Laiho, M., DeCaprio, J. A., Ludlow, J. W., Livingston, D. M. & Massagué, J. Growth inhibition by TGF- β linked to suppression of retinoblastoma protein phosphorylation. *Cell* **62**, 175–185 (1990).
- Derynck, R. & Zhang, Y. Intracellular signalling: the mad way to do it. *Curr. Biol.* **6**, 1226–1229 (1996).
- Massagué, J. TGF β signaling: receptors, transducers, and Mad proteins. *Cell* **85**, 947–950 (1996).
- Wrana, J. L. & Attisano, L. MAD-related proteins in TGF β signaling. *Trends Genet.* **12**, 493–496 (1996).
- Wu, R.-Y., Zhang, Y., Feng, X.-H. & Derynck, R. Heteromeric and homomeric interactions correlate with signaling activity and functional cooperativity of Smad3 and Smad4/DPC4. *Mol. Cell. Biol.* **17**, 2521–2528 (1997).
- Nakao, A. *et al.* TGF- β receptor-mediated signalling through Smad2, Smad3 and Smad4. *EMBO J.* **16**, 5353–5362 (1997).

- Wieser, R., Wrana, J. L. & Massagué, J. GS domain mutations that constitutively activate T β -R1, the downstream signaling component in the TGF- β receptor complex. *EMBO J.* **14**, 2199–2208 (1995).
- Yingling, J. M. *et al.* Tumor suppressor Smad4 is a transforming growth factor β -inducible DNA binding protein. *Mol. Cell. Biol.* **17**, 7019–7028 (1997).
- Tanaka, T. *et al.* Dual functions of the AML1/Evi-1 chimeric protein in the mechanism of leukemogenesis in t(3;21) leukemias. *Mol. Cell. Biol.* **15**, 2383–2392 (1995).
- Lagna, G., Hata, A., Hemmati, B. A. & Massagué, J. Partnership between DPC4 and SMAD proteins in TGF- β signalling pathways. *Nature* **383**, 832–836 (1996).
- Hata, A., S., L. R., Wotton, D., Lagna, G. & Massagué, J. Mutations increasing autoinhibition inactivate tumor suppressors Smad2 and Smad4. *Nature* **388**, 82–87 (1997).
- Takebe, Y. *et al.* SR α promoter: an efficient and versatile mammalian cDNA expression system composed of the simian virus 40 early promoter and the R-U5 segment of human T-cell leukemia virus type 1 long terminal repeat. *Mol. Cell. Biol.* **8**, 466–472 (1988).
- Kurokawa, M. *et al.* A conserved cysteine residue in the runt homology domain of AML1 is required for the DNA binding ability and the transforming activity on fibroblasts. *J. Biol. Chem.* **271**, 16870–16876 (1996).

Acknowledgements. We thank J. Massagué for Smad1–Flag and Smad4–HA; J. L. Wrana for Smad2–Flag and the pCMV5 vector; R. Derynck for Smad3–Flag and Smad4–Flag; X.-F. Wang for p15P113–Luc; K. Miyazono for p3TP–Lux and T β R1 TD; and K. Arai for pME18Sneo. This work was supported in part by Grants-in-Aid for Cancer Research from the Ministry of Health and Welfare and from the Ministry of Education, Science, and Culture of Japan.

Correspondence and requests for materials should be addressed to H.H. (e-mail: hhirai-ty@umin.u-tokyo.ac.jp).

DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI

Karen E. Flick*, Melissa S. Jurica*†, Raymond J. Monnat Jr‡ & Barry L. Stoddard*

*Fred Hutchinson Cancer Research Center and †Molecular and Cellular Biology Program of the University of Washington, 1100 Fairview Avenue North, Seattle, Washington 98109, USA

‡University of Washington, Department of Pathology, Box 357705, Seattle, Washington 98195, USA

Homing endonucleases are a diverse collection of proteins that are encoded by genes with mobile, self-splicing introns^{1–3}. They have also been identified in self-splicing inteins (protein introns)⁴. These enzymes promote the movement of the DNA sequences that encode them from one chromosome location to another; they do this by making a site-specific double-strand break at a target site in an allele that lacks the corresponding mobile intron³. The target sites recognized by these small endonucleases are generally long (14–44 base pairs). Four families of homing endonucleases have been identified, including the LAGLIDADG, the His–Cys box, the GIY–YIG and the H–N–H endonucleases¹. The first identified His–Cys box homing endonuclease was I-PpoI from the slime mould *Physarum polycephalum*^{5,6}. Its gene resides in one of only a few nuclear introns known to exhibit genetic mobility⁷. Here we report the structure of the I-PpoI homing endonuclease bound to homing-site DNA determined to 1.8 Å resolution. I-PpoI displays an elongated fold of dimensions 25 × 35 × 80 Å, with mixed α/β topology. Each I-PpoI monomer contains three anti-parallel β -sheets flanked by two long α -helices and a long carboxy-terminal tail, and is stabilized by two bound zinc ions 15 Å apart. The enzyme possesses a new zinc-bound fold and endonuclease active site. The structure has been determined in both uncleaved substrate and cleaved product complexes.

The structure of I-PpoI (Fig. 1) shows that the central interface of the enzyme dimer is small and highly solvated, with subunit contacts that bury only approximately 700 Å² of surface area (Fig. 2a). The C-terminal tails (residues 146–163) are domain-swapped and extend 34 Å across opposite monomers, burying an additional 900 Å² per subunit (Fig. 2b). In contrast to the LAGLIDADG homing endonucleases^{8,9}, the I-PpoI dimer interface is not tightly packed.

The first bound zinc ion (Fig. 3) is coordinated by a cluster of three Cys and one His ligands. One side chain (Cys 41) is contributed by an amino-terminal β -strand ($\beta 2$) and the remaining three side chains (Cys 10, Cys 105 and His 110) are from a short loop between $\beta 7$ and $\beta 8$. The binding for this metal is C-X₅₈-C-X₄-C-X₄-

H. The second zinc ion is coordinated by a short cluster of four side chains (Cys 125, Cys 132, His 134 and Cys 138). The motif for the second bound zinc ion is C-X₆-C-X₁-H-X₃-H, and resembles the zinc-binding sequence from retroviral nucleocapsid proteins¹⁰ but has a different structural fold. All eight residues involved in zinc

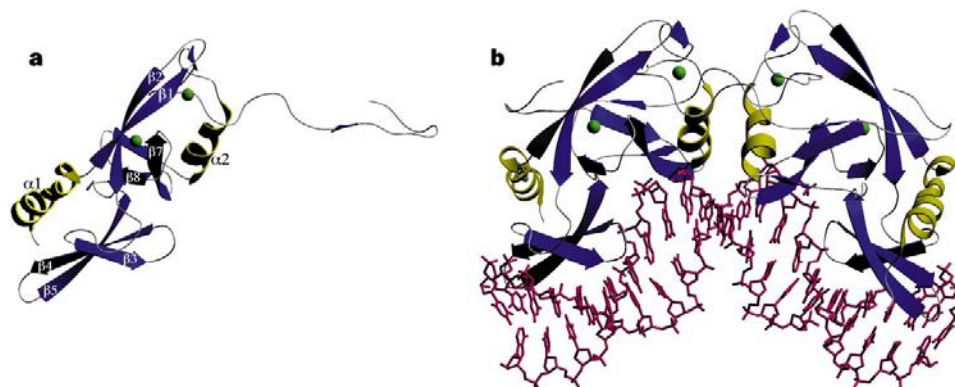


Figure 1 Tertiary fold and DNA complex of the *I-PpoI* endonuclease. **a**, The protein monomer is stabilized by two bound zinc ions and has a long C-terminal tail that extends around the surface of the second subunit in the dimer. **b**, The bound endonuclease dimer dramatically deforms the homing-site DNA, inducing

a sharp pair of kinks near the cleavage site opening the minor groove. The overall bend is approximately 50°. Figs 1, 2, 3 and 5 were made using the program SETOR²⁷.

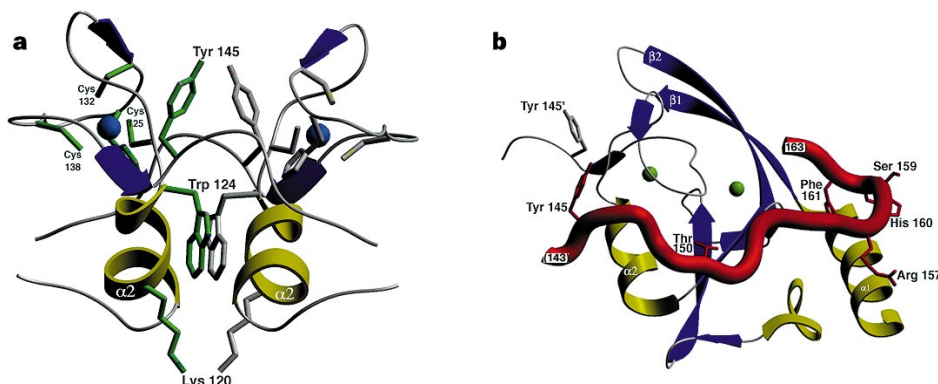


Figure 2 Structure of the dimer interface. **a**, The central dimer interface is composed primarily of Van der Waals contacts between Trp 124 and 124'. This interface buries a total of less than 700 Å² of protein surface. Lys 120 and 120' make direct contacts with thymine bases in the minor groove of the central cleavage site. **b**, The domain-swapped C-terminal tail (red) crosses the dimer

interface at Gly 146 and proceeds across the surface of the second subunit. Each arm results in an additional 900 Å² of buried surface area. Residues 150–153 form a short β -strand interaction with residues 111'–113'. Each of the C-terminal oxygens forms a hydrogen bond with His 40. Phe 160 is also buried in this interface.

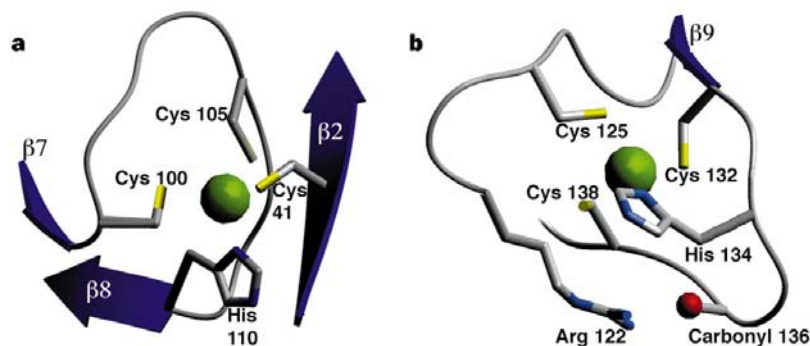


Figure 3 Zinc-binding motifs of *I-PpoI*. **a**, The N-terminal bound zinc is coordinated by Cys 41 from a long β -sheet and three additional residues in a short loop. **b**, The C-terminal zinc is coordinated by three cysteine and one

histidine residues contained within a short 14-residue sequence. This loop is flanked by a hydrogen bond between Arg 122 and the carboxyl oxygen of Val 136.

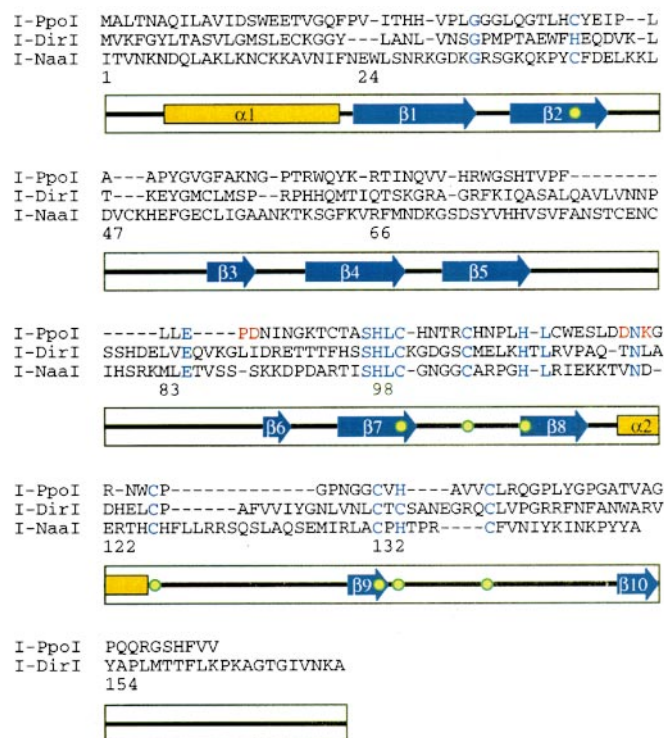


Figure 4 Sequence alignment of *I-PpoI* and other known nuclear His-Cys box homing endonucleases. The *I-PpoI* protein sequence is shown aligned with sequences from the homologues *I-DirI* and *I-NaaI* and its secondary structure elements. Residues that act as ligands to the two zinc atoms are denoted by circles. Conserved residues are blue; the residues in the previously noted 'endonuclease' PD...(E/D)X(K/R) motif in the *I-PpoI* sequence are red.

coordination are conserved as Cys or His residues among the three known His-Cys box homing endonucleases⁷ (Fig. 4). The conserved sequence Ser 97-His 98-Leu 99-Cys 100 contains one zinc-coordinating cysteine and an active-site histidine.

Previous structural studies have visualized three DNA-binding protein families containing two bound zinc ions per protein subunit: steroid receptors¹¹, the GAL4 transcription factor zinc-binding domain¹², and the RING finger motif¹³⁻¹⁵. Of these structures, only the steroid receptors coordinate two zinc atoms without sharing ligands between metals or using overlapping metal-binding sequences, as seen for *I-PpoI*¹¹. Rather than being directly involved in DNA binding, the *I-PpoI* zinc-binding motifs are critical primarily for structural stabilization of the protein core.

The primary sequence-specific contacts made by *I-PpoI* to homing-site DNA are between residues in the second β -sheet (β 3- β 4- β 5) of each enzyme monomer and base pairs 5-9 in the major groove of each half-site. Additional contacts are made in the centre of the complex within the minor groove and with several phosphate groups in the cleavage site. The *I-PpoI* homodimer strongly bends homing-site DNA for easier cleavage across the minor groove.

The antiparallel β -sheet that facilitates DNA binding forms a structure complementary to the major groove of B-form DNA (Fig. 5). This motif facilitates the recognition of extended DNA sequences by maintaining a spacing of 6.9 Å between alternate side chains and by adopting a curvature and twist complementary to the DNA helical axis¹⁶. This structure is similar to, but shorter than, the paired β -sheets formed by the *I-CreI* and *PI-SceI* homing endonucleases^{8,9}. The interactions formed in this interface represent both direct protein-DNA hydrogen bonds and water-mediated contacts, most of which are made by β 4, using every alternate side chain in

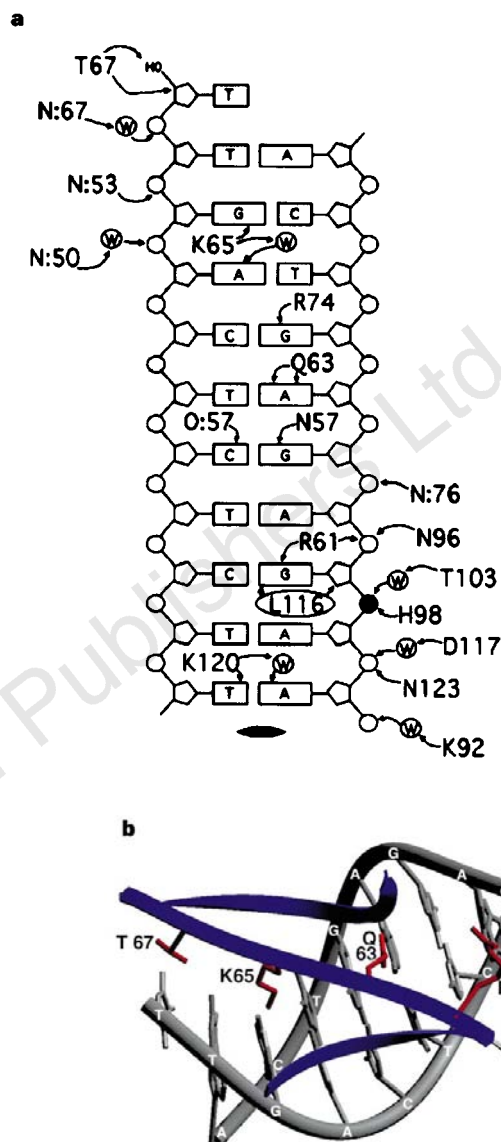


Figure 5 The endonuclease DNA-binding surface. **a**, Schematic diagram of contacts between an endonuclease monomer and a half-homing site. Residues 61, 63, 65 and 67 are all from the same face of a single β -strand and form a series of contacts down the major groove. **b**, The DNA-binding interactions of the *I-PpoI* β -ribbon in a homing half-site major groove. The alternating side chains from β 4 are shown.

its sequence. This β -strand begins with Arg 61, which contacts the phosphate backbone near the scissile phosphate. As the β -strand continues to follow the major groove towards the end of the homing site, every alternate residue contacts the DNA. Gln 63 makes direct bipartite contact to Ade +6. Genetic studies demonstrate that adenine is essential at this position (R.M., unpublished observations). Lys 65 contacts Ade -8 and Gua -9 through water-mediated interactions. Finally, the O γ of Thr 67 contacts the 5' terminus of the oligonucleotide. Although the side chain of Arg 66 in this central β -strand points away from the DNA, the long side chain wraps around the β -strand to make two water-mediated contacts with P-9 and P-10. A small number of additional contacts are made in this region by the two additional β -strands that flank β 4.

One base-specific contact in each half-site is made in the minor groove, where Lys 120 is bound to O2 of Thy 1 and makes a water-mediated contact to Ade 1. The A:T base pairs at these positions are strongly favoured, but can be reversed while still maintaining site

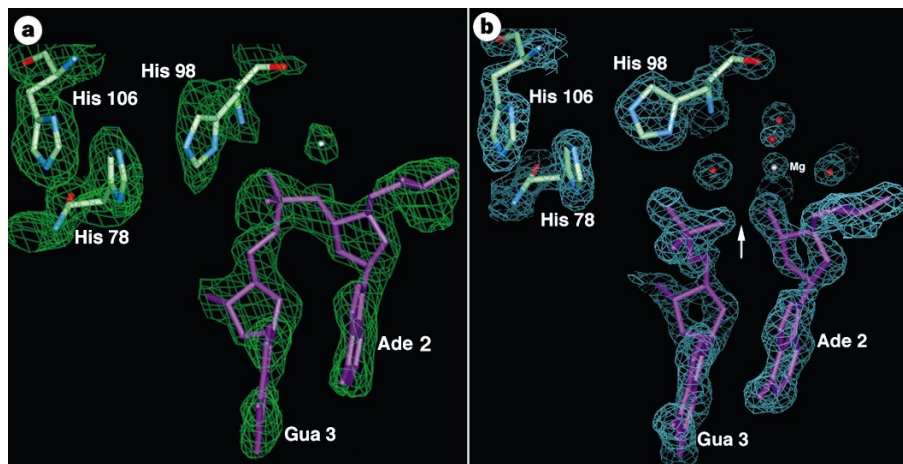


Figure 6 *I-PpoI* homing endonuclease active site. **a**, $2F_o - F_c$ simulated annealing omit map (2.2 Å) of the substrate complex contoured at 1σ , showing the density around the phosphodiester backbone between Ade 2 and Gua 3. The distance from the N δ nitrogen of His 98 to the scissile phosphate oxygen is 3 Å. **b**, A similar omit map (1.8 Å) of the cleaved product complex with a bound divalent cation. The 3' hydroxyl oxygen leaving group is 4 Å from the phosphorus atom. The octahedral coordination of the bound cation is clearly visible in the map. The cation is

directly coordinated by the 3' oxygen of the hydroxyl leaving group and by the side-chain oxygen of Asn 119. The asparagine ligand to the metal is not shown for clarity, but density for its oxygen atom is visible behind and below the magnesium ion. Arg 61, which makes a 2.9 Å contact to the cleaved phosphate group, is also not shown. Anomalous difference maps calculated from crystals soaked with 10 mM manganese verify the location of the bound metal ion. This figure was made using QUANTA²⁴.

recognition by the endonuclease (R.M., unpublished observations). The second base pair at either side of the central cleavage site shows a similar dependence on A:T content for endonuclease activity. The apparent exclusion of G:C base pairs in the cleavage site may reflect the presence of a specific contact in the minor groove that would be difficult to form in the presence of the bulkier guanine nucleotide.

Endonucleases such as *I-PpoI* that cleave across the minor groove of DNA (liberating 3' overhangs) must cleave two phosphate groups separated by only 10–12 Å in an undistorted B-form double helix. The homing endonucleases *I-CreI* and *PI-SceI* have evolved an elegant structural solution to this problem: the conserved LAGLI-DADG motif is used to generate a closely packed subunit or domain interface that appropriately positions active-site residues at the centre of the active site with minimal separation between them^{8,9}. *I-PpoI* uses a different strategy: it severely bends the homing-site DNA (Fig. 1), making the scissile phosphates more accessible for the two separate enzyme active sites.

At the point of the bend in each half-site, Leu 116 points into the minor groove and contacts Ade +2. Unlike the TATA-binding protein (TBP), which uses phenylalanine residues to intercalate between the bases of the TATA box^{17,18}, the *I-PpoI* leucines are not large enough to insert between the base pairs. Instead, they make edge-on contacts with the Ade 2, unstacking these bases from their immediate neighbours (Gua 3). This strategy, and the resulting distortion of the DNA, is similar to that observed for the complex of the purine repressor PurR to its regulatory binding site¹⁹. Mutation of Leu 116 greatly reduces the catalytic efficiency of *I-PpoI* (B. L. S., unpublished observations). The 55° bend across the central four base pairs of the homing site increases the average width of the minor groove in this region to 10 Å. The major groove across these same base pairs narrows to less than 5 Å, and the scissile phosphates are separated by 22 Å.

The strategy of DNA recognition and binding exhibited by the structure of the *I-PpoI* homing-site complex agrees well with previous biochemical and genetic analyses of this endonuclease⁵. *I-PpoI* requires at least 14 base pairs of its homing-site DNA sequence for efficient binding and cleavage, and it protects up to four base pairs to either side of the homing site in MPE-Fe(II) footprinting and protection analyses. A preference for purines flanking the homing site has been observed. This suggests that nonspecific contacts extending beyond the homing site may facil-

itate sequence-specific binding and cleavage. In the crystal structure, water-mediated base contacts and phosphate-backbone interactions extend beyond the direct protein–DNA contacts to generate an interface over 22 base pairs.

The structure of *I-PpoI* complexed to its DNA substrate was determined under three separate conditions to characterize the active site (Fig. 6). The enzyme was initially crystallized with low concentrations of magnesium. Electron density maps of this structure indicate that the scissile phosphate group has been cleaved, generating a product complex. The divalent cation is bound in the active site, where it displays six-fold octahedral coordination to four water molecules, one protein side chain and one DNA oxygen. Crystals grown under these conditions were soaked for 1 h in 10 mM manganese, and difference Fourier maps were used to confirm the location of the bound metal cation. Finally, a cocrystal structure was determined in which the metal-bound oxygen atom of the DNA substrate was substituted with sulphur. This strategy inhibits cleavage in the crystal. The resulting structure is of an uncleaved substrate complex.

In the substrate complex, the conserved His 98 is within hydrogen-bond distance of the scissile phosphate group. This residue is an obvious candidate to act as a Lewis acid and stabilize the negative charge on the pentacoordinate phosphate transition state. There is no indication that a water bound to an adjacent phosphate group might act as a nucleophile. We prefer a model in which a metal-bound water (visualized in the cleaved product complex) is deprotonated by a general base and attacks the phosphate. In this structure, there is a weak peak, near the position of the bound metal ion in the product complex, that may be either a water molecule or a low-occupancy magnesium site.

In the product complex, the phosphate group is cleaved with the distance between the 3' hydroxyl oxygen leaving group and the 5' phosphorus atom refining to approximately 4 Å. The coordinations of the bound magnesium and manganese ions, determined separately, are identical. The metal ion is coordinated to the 3' hydroxyl oxygen leaving group of the cleaved phosphate and to the side-chain oxygen of Asn 119 and has four visible coordinating water molecules, one of which directly contacts an oxygen atom of the cleaved 5' phosphate group. The metal–oxygen bond distances range from 2.0 to 2.4 Å and form octahedral coordination spheres around the metals in both cases. Apart from the active-site divalent cation and

Table 1 Crystallographic data

Crystal complex	Nat_1 Zn, Mg (cleaved)	NAT_ALS Zn, Mg (cleaved)	Nat_Cd Cd, Mg (cleaved)	Nat_Mn Zn, Mn (cleaved)	Thio_DNA Zn, thio-DNA (uncleaved)	Hg (derivative 1)	IdU (2)
Internal merging and scaling							
Resolution (Å)	2.7	1.8	2.1	2.8	1.9	2.7	2.7
Reflections measured	50,053	362,336	83,203	85,277	179,752	60,392	33,285
Unique reflections	19,611	71,892	30,702	16,111	50,210	17,300	14,471
Redundancy	2.55	5.04	2.71	5.29	3.58	3.49	2.30
Completeness (%)	91.9	99.8	89.8	95.2	96.3	92.4	77.9
Average $L/\sigma(I)$	15.8	19.1	15.3	12.3	18.4	14.5	12.2
R_{sym}	0.045	0.049	0.048	0.074	0.033	0.053	0.054
Mosaicity (°)	0.30	0.37	0.37	0.60	0.22	0.40	0.25
MIR phasing							
Resolution (Å)						3.0	3.0
Mean isomorphous difference						4.7	5.7
R_{iso} (%)						9.2	11.7
R_{cullis}						0.87	0.73
Phasing power						0.78	1.45
Refinement							
Resolution (Å)		1.8	2.1	2.8	1.9		
R-factor		22.6	21.2	23.2	20.1		
R_{free}		24.7	30.4	30.7	24.2		
No. of protein atoms		2,490	2,490	2,490	2,490		
No. of nucleic acid atoms		856	856	856	854		
Total solvent molecules		403	547	8	363		
R.m.s. bond lengths (Å)		0.011	0.010	0.010	0.013		
R.m.s. angles (°)		1.7	1.4	1.8	1.5		
Ramachandran distribution (% core, allowed, generous, disallowed)		91.8, 7.8, 0.4, 0.0	90.2, 9.4, 0.4, 0.0	84.8, 13.7, 1.2, 0.4	92.2, 7.0, 0.8, 0.0		
Mean B-factors (protein/DNA, solvent)		15.5, 20.7	31.9, 53.0	24.0, 47.0	18.6, 23.12		

its coordinating Asn residue, the other groups within potential chemical-interaction distance of the cleaved phosphate group are His 98 and Arg 61. The largest structural differences in the active site produced by cleavage are the movement of the scissile phosphate towards Arg 61 and a rotation of the His 98 side chain, which maintains its contact to the phosphate group.

I-PpoI displays a sequence that resembles a type II endonuclease signature (PD...(D/E)XK), formed by residues Pro 86–Asp 87... Asp 118–Asn 119–Lys 120. The residues identified in this motif are found in the crystal structure to not be involved in the enzyme active site. These residues are not conserved in a sequence alignment of the known His–Cys box endonucleases (Fig. 4). The endonuclease sequence signature identified in type II restriction endonucleases is only coincidentally present in I-PpoI (and probably in many other protein sequences), and does not serve an important role in endonuclease structure or function.

In summary, the long homing-site interface formed by the small I-PpoI endonuclease is facilitated both by an extended fold dependent on bound metals that largely substitute for a compact hydrophobic core, and by sparsely distributed protein–DNA contacts. Economy of structure appears to be an important feature of these proteins, which are probably limited in their overall size by the location of their open reading frames inside functional introns. Comparison of the structures of restriction and homing endonucleases provides a detailed view of how these proteins have independently evolved solutions to the problem of site-specific DNA recognition and cleavage. □

Methods

Crystals were of I-PpoI complexed with a 20-base-pair synthetic DNA duplex (Oligos, Etc.) containing a palindromic variant of the wild-type I-PpoI homing site²⁰. Crystals grown using this protocol contain a stable, cleaved product complex, with a single bound magnesium ion in each active site. The bound metal position was confirmed by soaking 10 mM manganese into crystals and calculating an anomalous difference Fourier map. Crystals of an uncleaved substrate complex were prepared in an identical manner, except that the hydroxyl oxygen of the 3' leaving group in the DNA oligonucleotide was substituted with a sulphur atom. All data, except set NAT_ALS (Table 1), were collected at 100 K on an RAXIS II-C area detector mounted on a Rigaku RU200

rotating anode X-ray generator (Molecular Structure Corporation). Data set NAT_ALS was collected on an ADSC 4 panel CCD area detector at beamline 5.0.2 at the Advanced Light Source (LBNL, Berkeley) using a wavelength of 1.0 Å. All RAXIS data were reduced using the DENZO/SCALEPACK crystallographic data reduction package²¹; the CCD data were reduced using MOSFLM 5.5 (ref. 22). One isomorphous derivative was prepared by synthetic incorporation of an iododeoxyuridine in the DNA duplex. The second was prepared by a 1-h soak of the native crystals in a 1 mM Hg(CN)₂ solution. All initial multiple isomorphous replacement phase calculations and phase refinements were done using the 'Native 1' data set in the CCP4 package²³ with phases calculated using the program MLPHARE. Phases calculated at 3.50 Å resolution using only these data yielded an interpretable electron density map that was further improved by solvent flattening, histogram matching and phase extension to 2.8 Å resolution. These phases were applied to the higher quality 'Cd' data set for model building, and extended to 2.1 Å resolution after the initial chain trace. Phase-combined maps were calculated using SFALL and SIGMAA to resolve ambiguous regions of the original DM electron density map. There is a single I-PpoI dimer and DNA duplex in the asymmetric unit of the crystal. Model building was carried out using QUANTA²⁴, followed by refinement using XPLOR²⁵. The higher-resolution NAT_ALS data set then became available for final refinement of the cleaved product complex to 1.8 Å resolution. The stereochemical quality of the protein model was examined throughout the refinement using PROCHECK²⁶. The final models contain no residues with unfavourable backbone dihedral angles with over 90% of the residues in the core region of the Ramachandran plot.

Received 19 January; accepted 8 April 1998.

- Belfort, M. & Roberts, R. Homing endonucleases—keeping the house in order. *Nucleic Acids Res.* **25**, 3379–3388 (1997).
- Belfort, M., Reaban, M. E., Coetzee, T. & Dalggaard, J. Z. Prokaryotic introns and inteins: a panoply of form and function. *J. Bacteriol.* **177**, 3897–3903 (1995).
- Belfort, M. & Perlman, P. S. Mechanisms of intron mobility. *J. Biol. Chem.* **270**, 30237–30240 (1995).
- Cooper, A. A. & Stevens, T. H. Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends Biochem. Sci.* **20**, 351–356 (1995).
- Ellison, E. L. & Vogt, V. M. Interaction of the intron-encoded mobility endonuclease I-PpoI with its target site. *Mol. Cell. Biol.* **13**, 7531–7539 (1993).
- Muscarella, D. E. & Vogt, V. M. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* **56**, 443–454 (1989).
- Johansen, S., Embley, T. M. & Willassen, N. P. A family of nuclear homing endonucleases. *Nucleic Acids Res.* **21**, 4405 (1993).
- Duan, X., Gimble, F. & Quijcho, F. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* **89**, 555–564 (1997).
- Heath, P. J., Stephens, K. M., Monnat, R. J. & Stoddard, B. L. The structure of I-Crel, a group I intron-encoded homing endonuclease. *Nature Struct. Biol.* **4**, 468–476 (1997).

10. South, T. L., Blake, P. R., Hare, D. R. & Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **30**, 6342–6349 (1991).
11. Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
12. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356**, 408–414 (1992).
13. Everett, R. D. *et al.* A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J. Mol. Biol.* **234**, 1038–1047 (1993).
14. Barlow, P. N., Luisi, B., Milner, A., Elliot, M. & Everett, R. Structure of the C₃HC₄ domain by ¹H-nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **237**, 201–211 (1994).
15. Borden, K. L. B. *et al.* The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J.* **14**, 1532–1541 (1995).
16. Phillips, S. E. V. The β-ribbon DNA recognition motif. *Ann. Rev. Biophys. Biomol. Struct.* **23**, 671–701 (1994).
17. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
18. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP–TATA-box complex. *Nature* **365**, 512–520 (1993).
19. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α-helices. *Science* **266**, 763–770 (1994).
20. Flick, K. E. *et al.* Crystallization and preliminary X-ray studies of I-Ppol: a nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. *Protein Sci.* **6**, 1–4 (1997).
21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
22. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1992).
23. CCP4 The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography (Daresbury Laboratory, Warrington, UK, 1979).
24. QUANTA96 X-ray Structure Analysis User's Reference (Molecular Simulations, San Diego, 1996).
25. Brünger, A. XPLOR version 3.1: A System for X-ray Crystallography and NMR (Yale Univ. Press, New Haven, CT, 1992).
26. Laskowski, R. J., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 383–390 (1993).
27. Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).

Acknowledgements. We thank D. McHugh, K. Stephens and J. D. Heath for initial subcloning, purification and crystallization studies; R. Strong, K. Zhang and B. Scott for advice during the crystallographic analysis; and the beamline staff at the Advanced Light Source (NLBL laboratories), beamline 5.0.2, particularly T. Earnest, for assistance. B.L.S. and R.J.M. are funded for this project by the NIH. K.E.F. was supported by a NIH training grant and the American Heart Association. M.S.J. was supported by an NSF fellowship and an NIH training grant.

Correspondence and requests for materials and coordinates should be addressed to B.L.S. (e-mail: bstoddar@fred.hcr.org). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession nos 1lpp, 1a73, 1a74).

corrections

Emergence of symbiosis in peptide self-replication through a hypercyclic network

David H. Lee, Kay Severin, Yohei Yokobayashi & M. Reza Ghadiri

Nature **390**, 591–594 (1997)

Hypercycles are based on second-order (or higher) autocatalysis and defined by two or more replicators that are connected by

another superimposed autocatalytic cycle. Our study describes a mutualistic relationship between two replicators, each catalysing the formation of the other, that are linked by a superimposed catalytic cycle. Although the kinetic data suggest the intermediary of higher-order species in the autocatalytic processes, the present system should not be referred to as an example of a minimal hypercycle in the absence of direct experimental evidence for the autocatalytic cross-coupling between replicators. □

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk, Rebecca A. Clayton, Jean-Francois Tomb, Owen White, Karen E. Nelson, Karen A. Ketchum, Robert J. Dodson, Michelle Gwinn, Erin K. Hickey, Jeremy D. Peterson, Delwood L. Richardson, Anthony R. Kerlavage, David E. Graham, Nikos C. Kyrpides, Robert D. Fleischmann, John Quackenbush, Norman H. Lee, Granger G. Sutton, Steven Gill, Ewen F. Kirkness, Brian A. Dougherty, Keith McKenney, Mark D. Adams, Brendan Loftus, Scott Peterson, Claudia I. Reich, Leslie K. McNeil, Jonathan H. Badger, Anna Glodek, Lixin Zhou, Ross Overbeek, Jeannine D. Gocayne, Janice F. Weidman, Lisa McDonald, Teresa Utterback, Matthew D. Cotton, Tracy Spriggs, Patricia Artiach, Brian P. Kaine, Sean M. Sykes, Paul W. Sadow, Kurt P. D'Andrea, Cheryl Bowman, Claire Fujii, Stacey A. Garland, Tanya M. Mason, Gary J. Olsen, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese & J. Craig Venter

Nature **390**, 364–370 (1997)

The pathway for sulphate reduction is incorrect as published: in Fig. 3 on page 367, adenylyl sulphate 3-phosphotransferase (*cysC*) is not needed in the pathway as outlined, as adenylyl sulphate reductase (*aprAB*) catalyses the first step in the reduction of adenylyl sulphate. The correct sequence of reactions is: sulphate is first activated to adenylyl sulphate, then reduced to sulphite and subsequently to sulphide. The enzymes catalysing these reactions are: sulphate adenylyltransferase (*sat*), adenylylsulphate reductase (*aprAB*), and sulphite reductase (*dsrABD*). We thank Jens-Dirk Schwenn for bringing this error to our attention. □

Refining your image acquisition

Kicking off with electron and atomic force microscopy, this section is rounded out by systems for light microscopy that include laser scanning fluorescence, improved automation, imaging software and other accessories.

Vacuum equipment

From Electron Microscopy Sciences

A complete line of vacuum equipment

This range of vacuum equipment includes stand-alone glow discharge units, etch facilities, film thickness monitors, carbon-coating units, sputter coating units, turbo sputter and carbon coaters, critical-point dryers, freeze dryers, turbo evaporators, plasma ashers and cryogenic preparation systems. All of the equipment is fully automatic and comes with an unconditional, two-year warranty on all parts and labour. Standard features include built-in rotary and tilt stages, built-in multi-specimen holders, built-in penning heads and gauges. The Electron Microscopy Sciences range is designed to accommodate multiple accessories and attachments.

Reader Enquiry No. 100

MAC Mode

From Molecular Imaging

The magnetic AC (MAC) mode for atomic force microscopes is now an option with this scanning probe microscopy (SPM) line

MAC Mode vertically oscillates the AFM probe using magnetic cantilevers. According to the manufacturer, it is the best means of achieving high resolution with the AC technique, and is well suited for imaging delicate samples in fluids. The key to high-resolution, *in situ* imaging, says the manufacturer, is driving the cantilever directly, rather than exciting the fluid in order to drive the tip. Biological samples can now be imaged *in vitro* in biological buffers (37 °C). MAC Mode is compatible with the company's temperature and environmental control, as well as with Digital Instruments, Park, Topometrix and RHK controllers.

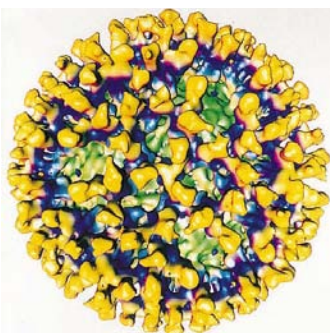
Reader Enquiry No. 101

CT3500 Cryotrans

From Oxford Instruments

Three-dimensional, macromolecular reconstruction from ultra-high-resolution TEM results

The CT3500 Cryotrans holder and transfer system for transmission electron microscopy (TEM) has found a new role, enabling three-dimensional reconstruction of biological core-like particles, such as those associated with the bluetongue virus (BTV). The essence of the new technique is to freeze the molecules rapidly within a thin film of amorphous ice and record a series of images of the molecule from different orientations.



Oxford Instruments helps rebuilding molecules.

The images are then digitized and the different orientations are recombined with Fourier analysis. This is said to result in a three-dimensional image that provides a resolution of 0.34 nm in all directions. The system provides stable, low-temperature performance even at high tilt angles within the TEM. Its design is said to provide high specification and ease of use so that ultra-high resolution can be routinely achieved in any laboratory. It has a temperature controller and a heat exchanger, and offers frost-free sample loading and transfer for both the sample and the holder. However, only the specimen area is cooled, and not the holder tip, a method that is said to result in low drift. There is also an advanced Dewar, which is used to ensure that the cooling liquid does not boil, thus reducing the vibration associated with boiling liquid nitrogen and improving resolution.

Reader Enquiry No. 102

S-3000 series

From Hitachi

A new series of entry-level, PC-controlled scanning electron microscopes (SEMs)

For conventional high-vacuum SEM there is the S-3000H and for variable pressure SEM there is the S-3000N. The S-3000 series has an accelerating voltage range of 0.3–30 kV and uses a tungsten emitter with an automatic, dual-bias function. This is said to give improved resolution at high accelerating voltages, and the dual-bias system is said to boost the emission current by tenfold at voltages lower than 5 kV, providing improved resolution and signal to noise. There is a high-resolution frame store of 1,280 × 960 pixels and comprehensive image-processing facilities, which include recursive filtering, frame integration and four-quadrant, split-screen image display. The system uses a Windows95 user interface with a smart mouse to reduce mouse movement and clicking; for versatility, a set

of rotary controls is provided. Also offered are automatic functions for focus, stigmator, brightness and contrast, and start up. Moreover, brightness, contrast, focus and astigmatism can be adjusted using the mouse. With a six-inch chamber and a choice of five specimen stages, the S-3000 series can accommodate a wide range of different samples, and the S-3000N variable pressure instrument offers the added versatility of being able to examine hydrated specimens in their natural state.

Reader Enquiry No. 103

Lumina

From TopoMetrix

A system that integrates near-field optical, atomic force and inverted optical microscopy

This integrated design combines optical microscopy with near-field optical (NSOM) and atomic force microscopy (AFM). All three techniques can be applied to the same sample in a matter of minutes with this integrated instrument. A new brochure that illustrates the benefits of combining these techniques highlights the using polytene chromosomes as an example of its application. Key design features include tuning-fork NSOM probes, TrueMetrix linearized scanners and auxiliary device compatibility. The brochure provides current specifications to give the prospective user a concise overview of the system. Moreover, examples and other system products and application developments can be found on the company's website: <topometrix.com>

Reader Enquiry No. 104

Multimicroscopy

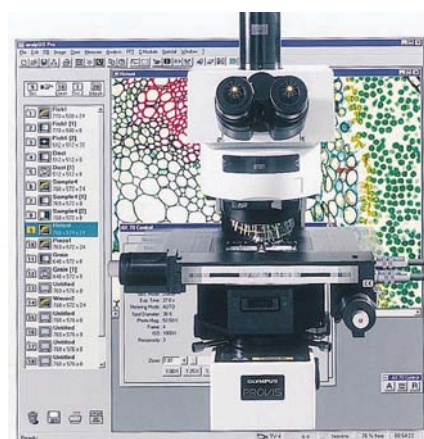
Rotary mechanical stage

From Carl Zeiss

This new rotating stage has a ceramic coating and should prove useful in routine laboratory work with the Axiolab microscope

This stage increases the flexibility of this microscope range, particularly where specimens need to be rotated for photomicrography or video microscopy applications. Rotation through 210° can be achieved and the stage can then be locked *in situ* at any angle. To allow precise final positioning, there are coaxial controls for the *x-y* movement of the specimen and Vernier scales for basic location. For improved longevity, the surface of the stage has been hard anodized and then ceramically coated to reduce scratching of the anodized surface.

Reader Enquiry No. 105



Keeping CALM with the AX70 microscope.

CALM

From Olympus

Computer-aided light microscopy (CALM)

This system combines the AX70 Provis automated research microscope with a motorized scanning stage and the PC-based analySIS-Pro image-analysis system. This image analysis system operates in a Windows environment, and has been extended with plug-in modules, which allow the control of the motorized system components. It also offers new automated capabilities through the use of a macro recorder, which allows all the steps of a complex microscopic analysis to be performed automatically. The system should prove useful in all fields of biology and materials science where high numbers of specimens have to be analysed or routine analysis standardized.

Reader Enquiry No. 106

OZ confocal microscope

From Noran Instruments

The O₂ UNIX workstation from Silicon Graphics is included with this confocal laser scanning microscope

The O₂ workstation supports the video rate and faster scanning capabilities of the OZ system, using rapid digital-storage capabilities.



Pure O₂ with Noran's OZ confocal microscope.

The O₂ processor enables rapid ratiometric analysis of image series and allows interactive rotation of three-dimensional reconstructions from large stacks of confocal images. The workstation integrates video output as a standard feature, so that presentation videos for lectures and talks can be created easily.

Reader Enquiry No. 107

Imaging software

MetaMorph

From Universal Imaging

New software for fluorescence microscopy

This software offers three new software modules for multicolour immunofluorescence microscopy work. Researchers can acquire multiple wavelengths by choosing settings for both a camera (exposure, averaging, binning and region of image) and illumination devices (filter wheels, shutters, monochromators), which are specific to each fluorophore in the sample. Images can then be collected in one, several or all wavelengths with just the push of a button. The camera exposure settings may also be set automatically to optimize the dynamic range of each image. There is also a function that allows users to overlay fluorescent images, adding a new tool for combining fluorescent and transmitted light images to this system. With this tool, the researcher can combine the images from two fluorescent probes or proteins with any monochrome image into a single-colour image. Using new colour-balancing algorithms, a choice of the transparency and colour of the fluorescent images can be made so as not to obscure details in the monochrome transmitted light images. There is also 'co-localize', a tool that analyses two images representing two fluorescent proteins or fluorescent probes. It quantifies the overlap and non-overlap between the two fluorescent proteins or probes as a measure of co-localization. As with all MetaMorph measurements, results can be stored as text files or transferred directly to a spreadsheet.

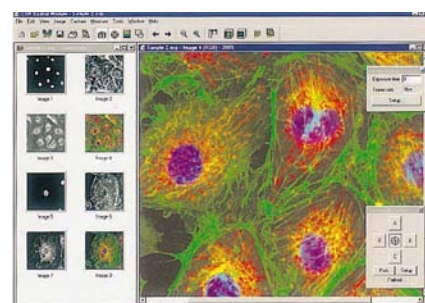
Reader Enquiry No. 108

UltraPlus

From LSR AstroCam

A camera control and image analysis software interface for the company's range of digital CCD cameras

UltraPlus is designed to be a comprehensive, high-performance low-light-level image acquisition and analysis software interface for the UltraPix range of 12- and 16-bit cooled CCD cameras. The Windows95 interface comprises two user friendly modules: camera control and spatial. These interactive modules provide a flexible working environment for both static and dynamic fluorescence, luminescence and bright-field imaging. The interface should prove useful in biological and medical applications, as well as in the



LSR AstroCam: seeing in the dark.

physical sciences, industrial inspection and non-destructive testing. The spatial module is used for grabbing multiple wavelength images or groups of images, which are then displayed in the thumbnail view gallery. Morphometric measurements and densitometric quantification can be applied to up to 100 user-defined regions within a chosen image. The camera-control module allows on-chip, asymmetric binning and sub-array selection, with image acquisition rates of up to 1,500 images per second. Automated time-lapse studies, detailed image statistics and on-line diagnostics are also provided. Images are can be processed using background subtraction, flat-field correction and thresholding. According to the manufacturer, a seamless interface is also provided between the UltraPix range of cameras and third-party image processing and analysis packages, including Optimas and Image Pro Plus.

Reader Enquiry No. 109

Fluoro-Pro software

From Media Cybernetics

An automated method of acquiring and processing multichannel fluorescent data

Fluoro-Pro is designed to simplify the acquisition of multicolour, fluorescent samples, such as those combining fluorescence *in situ* hybridization and green fluorescent protein. The software provides automatic exposure support for both analog and digital cameras. Image capture is said to be fast and repeatable, and the automation capabilities help to minimize photobleaching. Captured images can then be archived into the Image-Pro Plus database. When used in conjunction with Scope-Pro, the Fluoro-Pro interface also allows users to capture images with user-defined filter sets, microscope settings and stage positions. Fluoro-Pro is supported by Image-Pro's internal programming language. Moreover, repetitive functions and movements can be automated by recording them in macros, which can be executed in Image-Pro or through customized interfaces in Visual Basic. By integrating control of the stage with a macro program, users can incorporate some 200 Image-Pro analysis and processing routines into a fully automated image capture, analysis, and data collection process.

Reader Enquiry No. 110

Accessories

SPOT-2

From Diagnostic Instruments

A new camera with a dual-position colour filter that captures colour or high-sensitivity 12-bit monochrome images

This cooled CCD, colour digital camera has been designed specifically for microscopes. SPOT-2 has been updated from its predecessor by incorporating a colour and IR filter mounted on a sliding lever mechanism. When the colour and IR filter are removed from the light path, the camera sensitivity increases substantially. This is said to result in much shorter exposure times because the colour filter absorbs a significant amount of light, which should prove useful for low-light-level microscopy techniques, such as fluorescence. Removal of the IR filter allows higher wavelengths of light to reach the CCD. This is important for capturing light emission from Cy5 and Cy7 fluorescence dyes. The new SPOT-2 and original SPOT camera will come with new stand-alone software, which will allow the user to select easily among 8-bit monochrome, 12-bit monochrome, 24-bit colour, or 36-bit colour image acquisition. The software will also have binning modes (2×2, 3×3 and 4×4), region-of-interest selection, faster low-light-level focusing, a measurement module for calculating lengths, areas, angles and curves, and a merge image function.

Reader Enquiry No. 111

CFI IR objective lens

From Nikon

A new infrared (IR) objective CFI plan apochromat ×100 oil lens

Infrared light is well suited for the examination of thicker specimens and can be less damaging to cells examined under the microscope. With only a single coating on the inside of the objective lens, improvements in transmission ratios of 60–70 per cent at 900–1,000 nm are possible when compared with the normal CFI plan apochromat ×100 oil lens. Designed for use with the Eclipse E600FN, this new IR objective lens should be useful with laser tweezers and laser cutters, and multi-photon, IR-fluorescence, OEM and bright-field observations. According to the manufacturer, the operator can easily change objectives in the close confines of a Petri dish or deep chamber, without disturbing the specimen. The Eclipse E600FN, was specifically created for electrophysiology studies and the observation and micromanipulation of semi-opaque specimens and tissue cultures. With a focusing nose-piece, this microscope provides space for up to five micromanipulators.

Reader Enquiry No. 112

Coated microscope slides

From Xenopore

Glass microscope slides with ligands covalently attached: streptavidin, metal chelates, protein A, protein G, con A and glutathione

According to the manufacturer, the clarity of the slides is unaffected by binding these ligands. Slides are also available with covalent attachment sites for the immobilization of cells and antibodies, and can be used for cell panning, *in situ* hybridization, and other applications. The same surface coatings are available on glass cover slips.

Reader Enquiry No. 113

Treasury NT

From Synoptics

An image database for Windows NT

A 32-bit version of the Treasury database management system has been released for use with Windows NT and Windows95. It is also available as an upgrade to existing users. Treasury NT is said to operate faster than standard Windows versions and has been designed to deal even more efficiently with larger (higher resolution) images. Two new features have been introduced, including a mode of operation that stores images on a suitable bulk-storage medium, accessing them only when the appropriate data record is recalled. Treasury NT also offers a

'bulk move' facility, which will allow groups of images to be moved from one location to another. In addition a 'read only' facility has been introduced for network use. This allows database records and images to be accessed from anywhere on the network, but not changed.

Reader Enquiry No. 114

ImageStation 440CF

From Kodak and NEN Life Science

A compact benchtop detection and analysis system for chemiluminescence, fluorescence, chromogen detection and densitometry

Designed to eliminate the need for a dark room, this system performs quantitative analysis on a wide range of sample types, including membranes, gels, and X-ray film. The system uses an electronically cooled, full-frame-capture CCD camera to visualize flat-format samples. It is said to have a sensitivity on the order of picomoles to femtomoles per band on gels and blots. Moreover, it has ×6 zoom lens to cover the range from 20×25 cm at full field to 3.3×4.2 cm at full zoom, providing very high resolution.

Reader Enquiry No. 115

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

ANTIBODIES

For Research & Routine Histology

Inflammation – Angiogenesis – Tumors
Leukocytes – Endothelial Cells
Pathological Markers

Custom Monoclonals and Polyclonals

BMA Biomedicals Ltd.
(800) 634-3183 (US)
Intl: +41 61 811 62 22
+41 61 811 60 06 (fax)
info@bma.ch

READER ENQUIRY NO. 28

Detect residual DNA with Wako's

DNA Extractor Kit

or Isolate Genomic DNA with

DNA Extractor WB Kit

**Both use the Nal Method
No Phenol No Chloroform**

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 39

NEW ANTIBODIES FOR UBIQUITIN AND PROTEASOME RESEARCH FROM AFFINITI

20S Proteasome α- & β-subunits: HC2, HC3, HC8, HC9, iota, XAPC7, zeta, HC10, HN3, Z, MECL1, LMP2, LMP7; 19S regulator ATPase & non-ATPase subunits; 11S regulator subunits: PA28α, β and γ; **Signalosome** subunits; PGP9.5; E2; Ub-conjugates

For data sheets or more information on AFFINITI's extensive range of products for ubiquitin and proteasome research visit our website at:

HTTP://WWW.PROTEASOME.COM

AFFINITI Research Products Ltd., Mamhead Castle, EXETER, EX6 8HD, UK
Tel. +44/0 1626 891010; Fax: +44/0 1626 891090; URL: <http://www.affiniti-res.com>

READER ENQUIRY NO. 38